



UNIVERSIDADE FEDERAL DE CAMPINA GRANDE
CENTRO DE ENGENHARIA ELÉTRICA E INFORMÁTICA
COORDENAÇÃO DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

Aline Marques de Moraes

**MAPEAMENTO DA DIVERSIDADE DE CRIADORES DE CONTEÚDO
GEOLOCALIZADO EM REDES GEOSOCIAIS EM CIDADES
BRASILEIRAS**

Campina Grande, Paraíba, Brasil

2018

Aline Marques de Moraes

**MAPEAMENTO DA DIVERSIDADE DE CRIADORES DE CONTEÚDO
GEOLOCALIZADO EM REDES GEOSOCIAIS EM CIDADES
BRASILEIRAS**

Tese apresentada ao Programa de Pós-Graduação
em Ciência da Computação da Universidade
Federal de Campina Grande, como requisito
parcial para obtenção do título de Doutora em
Ciência da Computação.

Orientador: Prof. Dr. Nazareno Ferreira Andrade
Área de concentração: Sistemas de Computação

Campina Grande, Paraíba, Brasil

2018

FICHA CATALOGRÁFICA ELABORADA PELA BIBLIOTECA CENTRAL DA UFCG

M827m Morais, Aline Marques de.

Mapeamento da diversidade de criadores de conteúdo geolocalizado em redes geossociais em cidades brasileiras / Aline Marques de Morais. – Campina Grande, 2018.
171 f.

Tese (Doutorado em Ciência da Computação) – Universidade Federal de Campina Grande, Centro de Engenharia Elétrica e Informática, 2018. "Orientação: Prof. Dr. Nazareno Ferreira Andrade".
Referências.

1. Sistemas de Computação. 2. Cidades Brasileiras - Redes Geossociais. 3. Criadores de Conteúdo. I. Andrade, Nazareno Ferreira. II. Título.

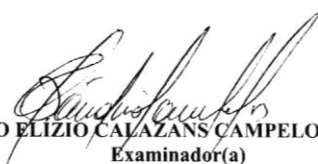
CDU 004.78 (043)

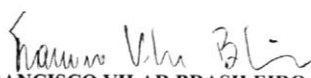
**"MAPEAMENTO DA DIVERSIDADE DE CRIADORES DE CONTEÚDO
GEOLOCALIZADO EM REDES GEOSOCIAIS EM CIDADES BRASILEIRAS"**

ALINE MARQUES DE MORAIS


TESE APROVADA EM 29/01/2018


NAZARENO FERREIRA DE ANDRADE, Dr., UFCG
Orientador(a)


CLÁUDIO ELÍZIO CALAZANS CAMPELO, Ph.D., UFCG
Examinador(a)


FRANCISCO VILAR BRASILEIRO, Ph.D., UFCG
Examinador(a)

FERNANDO MARQUES FIGUEIRA FILHO, Dr., UFRN
Examinador(a)


MARCUS WILLIAMS AQUINO DE CARVALHO, Dr., UFPB
Examinador(a)

CAMPINA GRANDE - PB

Agradecimentos

Inicialmente, gostaria de agradecer aos meus pais, Alberto Marques e Maria Nicolau, por todo o apoio durante essa longa jornada. Obrigada por toda serenidade transmitida nos dias de estresse. Quero agradecer também às minhas melhores amigas e irmãs Alana e Sarah, por toda a cumplicidade dessa temporada. Outra pessoa importante é meu marido Eduardo por toda paciência, especialmente quando estávamos longe um do outro, e compreensão que estava fazendo aquilo que realmente gosto, que é a pesquisa. Por fim, gostaria de agradecer ao meu orientador Nazareno Andrade pela parceria desses últimos anos e pelo exemplo de docência. A todos aqueles, que de alguma forma me ajudaram nessa temporada da tese, o meu mais sincero obrigada.

Resumo

Redes Geossociais (RGs) representam um tipo de sistema colaborativo com informação geolocalizada encapsulada no conteúdo compartilhado. Essa formatação possibilita a interseção de informações presentes tanto no mundo real, quanto no virtual. Atualmente, a popularização das RGs já é algo concreto pelo mundo. Uma das razões para esse fenômeno é o suporte das RGs à tomada de decisão dos colaboradores sobre lugares. Além do mais, esses sistemas representam importantes ferramentas de monitoramento urbano, baseadas no fluxo de participação dos usuários. Entretanto, a presente tese detecta como problema de pesquisa o modo pelo qual as RGs tratam homogeneamente os seus participantes. Algumas consequências desse tratamento padronizado são um menor aproveitamento das potencialidades colaborativas dos usuários e uma menor eficiência na tomada de decisão dos participantes. Assim, este trabalho considera necessário a definição de parâmetros que caracterizem a diversidade de um tipo específico de participante: os criadores de conteúdo. Eles representam uma parcela dos usuários que efetivamente colaboram com as RGs. Trabalhos anteriores já investigaram aspectos da diversidade de criadores de conteúdo, em RGs usadas em diferentes regiões do mundo. Mas, nenhum deles analisou a diversidade dos criadores de conteúdo em RGs usadas em cidades brasileiras. Portanto, o presente trabalho realiza um mapeamento em larga escala da diversidade dos criadores de conteúdo geolocalizado em cidades brasileiras. Esse mapeamento é dividido em três abordagens. A primeira delas é a investigação da diversidade dos criadores, em relação à experiência dos contribuidores sobre os locais. A segunda abordagem é sobre a investigação da diversidade baseada nos padrões de comportamento ao contribuir. E, por fim, a última etapa do mapeamento se refere à diversidade dos participantes em relação ao comportamento espacial. Os resultados dos mapeamentos mostram que a diversidade dos criadores de conteúdo está presente em RGs usadas nas cidades brasileiras. Além do mais, a utilização dos resultados das abordagens é imprescindível para a concepção de RGs sensíveis aos tipos dos criadores e para mecanismos de recomendação de lugares.

Palavras-chave: Redes Geossociais, diversidade, criadores de conteúdo, cidades brasileiras.

Abstract

Geo-Social Networks (GSNs) are a type of collaborative systems with geolocated information as their shared content. This specific content allows the linking between real and virtual worlds. Nowadays, the popularization of GSNs is something real in world. One reason of that phenomenon is because GSNs support their users during the making decision process about places. Moreover, GSNs are important tools for urban monitoring, based on flow of users' participation. The present thesis detects as problem the way how GSNs treat their collaborators homogeneously. As consequence, GSNs do not use all collaboratives potentialities of their users and are not efficient during making decision process. We reinforce the necessity of definition of parameters about diversity of a specific user: the content creator. He presents a piece of users who shared geolocated information in GSNs. Previous studies did investigate aspects about creators' diversity on GSNs over the world. But, none of them analyzed this diversity among content creators of GSNs in Brazilian cities. The present thesis does a large-scale mapping about geolocated content creators in cities of Brazil. This mapping is divided in three approaches. The former is the investigation of diversity, based on collaborators' experiences about places. The next one is about diversity of content creators, according to their collaborative behaviors. The last approach is about diversity analysis, face to spatial patterns among GSNs participants. The results find out the diversity among geolocated content creators in Brazilian cities. Moreover, the results can be applied in GSNs designing and mechanism of places' recommendation.

Keywords: *Geo-Social Networks, Diversity, Content Creators, Brazilian Cities.*

Lista de Figuras

Figura 1. Trabalhos coletados e selecionados pela expressão de busca, por repositório	33
Figura 2. Metodologias presentes nos trabalhos sobre a diversidade dos criadores de conteúdo geolocalizado.....	38
Figura 3. Categorias de locais de acordo com a proporção de turistas nos lugares.....	50
Figura 4. Categorias de dicas compartilhadas pelos grupos de criadores de conteúdo	52
Figura 5. Proporção de criadores de conteúdo por cidade brasileira.....	66
Figura 6. Histograma das métricas da RG sobre criadores de conteúdo. Todas as métricas estão normalizadas pela técnica do z-score para apresentarem tamanhos comparáveis.....	67
Figura 7. Resultados da correlação de Kendall por pares de métrica.....	68
Figura 8. Gráfico sobre a técnica do cotovelo que representa o número de <i>clusters</i> pelas distâncias de <i>cluster</i> média.....	69
Figura 9. Análise de acordo com o método de <i>Silhouette</i>	69
Figura 10. Comportamento das métricas colaborativas por grupo de criador de conteúdo	70
Figura 11. Distribuição dos gêneros por grupo de criador de conteúdo nas RGs.....	74
Figura 12. Intensidade da participação de criadores pouco participativos (A), criadores medianos (B), criadores interessados por locais populares (C) e produtores (D) no Brasil.....	75
Figura 13. Percentual dos tipos de criadores em cada cidade analisada do Brasil.....	76
Figura 14. Distribuição dos grupos de criadores de conteúdo nas cidades observadas do Brasil.....	110
Figura 15. Correlação de Kendall da participação colaborativa dos criadores pouco participativos e as características urbanas das cidades.....	111
Figura 16. Correlação de Kendall dos padrões de comportamento ao contribuir dos criadores medianos e as características urbanas das cidades.....	112
Figura 17. Correlação de Kendall dos padrões de comportamento ao contribuir dos criadores interessados em locais populares e as características urbanas das cidades	113
Figura 18. Correlação de Kendall dos padrões de comportamento ao contribuir dos criadores produtores e as características urbanas das cidades.....	114

Lista de Quadros

Quadro 1. Áreas da cidade contempladas pelas RGs.....	36
Quadro 2. Resultados do teste de Wilcoxon.....	59
Quadro 3. Resultados do teste de Wilcoxon.....	60
Quadro 4. Comportamento espacial das colaborações dos criadores de conteúdo, de acordo com a autocorrelação espacial	85
Quadro 5. RGs selecionadas em trabalhos científicos com impacto nas cidades.....	105
Quadro 6. Linha do tempo das atividades por bimestre do ano	109
Quadro 7. Distribuição da participação dos colaboradores nas áreas urbanas de cada cidade brasileira selecionada.....	128
Quadro 8. Resultados da análise de clusters e <i>outliers</i> , sob enfoque do índice de Anselin Local Moran	141
Quadro 9. Resultados da análise de <i>Hotspots</i> e <i>Coldspots</i> das cidades brasileiras selecionadas	154
Quadro 10. Características urbanas que descrevem cada setor censitário das cidades selecionadas.....	167
Quadro 11. Resumo da análise de correlação de Kendall entre a quantidade de colaborações e características urbanas, (+) indica correlação positiva e (-) correlação negativa	169

Lista de Tabelas

Tabela 1. Resultados do modelo de regressão binomial negativa binomial negativa	55
Tabela 2. Resultados do modelo de regressão binomial negativa.....	58
Tabela 3. Resultados da autocorrelação espacial em cidades do Sudeste	115
Tabela 4. Resultados da autocorrelação espacial em cidades do Norte.....	120
Tabela 5. Resultados da autocorrelação espacial em cidades do Sul	121
Tabela 6. Resultados da autocorrelação espacial em cidades do Centro-Oeste	123
Tabela 7. Resultados da autocorrelação espacial em cidades do Nordeste	124

Lista de Siglas

EUA – Estados Unidos da América

GSN – *Geo-Social Network*

IBGE – Instituto Brasileiro de Geografia e Estatística

IDH – Índice de Desenvolvimento Humano

IMMASE – *Intelligent Mobile Multimedia Application for the Support of the Elderly*

ONU – Organização das Nações Unidas

POI – *Point of Interest*

RDH – Relatório de Desenvolvimento Humano

RG – Rede Geossocial

TF-IDF – *Term Frequency - Inverse Document Frequency*

Sumário

1 INTRODUÇÃO	27
1.1. Problema de pesquisa	28
1.2. Motivação.....	29
1.3. Objetivos	29
1.4. Contribuições	30
1.5. Sumário da tese.....	31
2 ESTADO DA ARTE.....	32
2.1. RGs nas cidades	32
2.2. Diversidade dos criadores de conteúdo geolocalizado	37
2.3. Métricas da diversidade de criadores de conteúdo geolocalizado.....	40
2.4. Discussão do embasamento.....	42
3 DADOS USADOS PARA A ANÁLISE DA DIVERSIDADE DOS CRIADORES DE CONTEÚDO	43
4 DIVERSIDADE BASEADA NA EXPERIÊNCIA	47
4.1. Relevância do conteúdo geolocalizado	47
4.2. Metodologia	48
4.3. Locais de turistas vs. locais de residentes.....	50
4.4. Visão geral dos dados.....	50
4.5. Quais fatores afetam a relevância do conteúdo compartilhado pelos grupos de criadores de conteúdo, durante o evento turístico?	53
4.6. Quão relevantes são as anotações criadas pelos grupos de colaboradores para a comunidade da RG, durante o evento turístico?	57
4.7. Quão relevantes são as anotações produzidas pelos grupos de colaboradores para a comunidade da RG, depois do evento turístico?	58
4.8. Discussão dos resultados	60
5 DIVERSIDADE BASEADA NOS PADRÕES DE COMPORTAMENTO AO CONTRIBUIR	62
5.1. Aspectos metodológicos	62
5.2. Padrões de comportamento ao contribuir dos criadores de conteúdo	63
5.3. Visão geral dos dados.....	65
5.4. Definição da quantidade de grupos de colaboradores para clusterização	68
5.5. Padrões da participação colaborativa dos criadores de conteúdo de RG em cidades brasileiras	70
5.5.1. Criadores de conteúdo pouco participativos	71

5.5.2. Criadores de conteúdo mediano.....	71
5.5.3. Criadores interessados por locais populares	71
5.5.4. Criadores produtores	72
5.6. Impressões sobre os grupos de criadores de conteúdo	73
5.7. Discussões dos resultados.....	78
5.8. Conclusões do mapeamento.....	79
6 DIVERSIDADE BASEADA NO COMPORTAMENTO ESPACIAL.....	81
6.1. Aspectos metodológicos	81
6.2. Resultados obtidos.....	83
6.2.1. Diversidade do comportamento espacial de criadores de conteúdo, por região das cidades brasileiras	83
6.2.2. Diversidade do comportamento espacial de criadores de conteúdo, baseado no IDH..	86
6.2.3. Diversidade do comportamento espacial de criadores de conteúdo, baseado no gênero majoritário da cidade	87
6.3. Discussão do mapeamento	88
7 CONCLUSÃO	90
7.1. Sumário	90
7.2. Contribuições para a pesquisa em RGs	91
7.3. Contribuições para o <i>design</i> de RGs.....	91
7.4. Beneficiados pela pesquisa	92
7.5. Publicações.....	93
7.6. Trabalhos futuros	94
7.7. Considerações finais.....	94
BIBLIOGRAFIA.....	96
APÊNDICES	105
Apêndice A – Descrição das RGs selecionadas na busca sistemática	105
Apêndice B – Linha do tempo das atividades descritas	109
Apêndice C – Participação dos perfis de criadores por cidade brasileira selecionada	110
Apêndice D – Correlação entre características da participação colaborativa de criadores e as características urbanas das cidades brasileiras selecionadas.....	111
Apêndice E – Resultados da autocorrelação espacial, sob o enfoque do índice de Moran.....	115
Apêndice F – Distribuição da participação de criadores de conteúdo geolocalizado em cada cidade brasileira selecionada.....	128

Apêndice G – Resultados da análise de <i>clusters</i> e <i>outliers</i> , de acordo com o índice de Anselin Local Moran, por cidade de cada região	141
Apêndice H – Resultados da análise de áreas muito frequentadas (<i>Hotspots</i>), de acordo com a estatística de Getis-Ord G_i^*	154
Apêndice I – Fatores urbanos escolhidos para a análise de correlação.....	167
Apêndice J – Fatores urbanos que afetam a intensidade de colaborações nos setores censitários das cidades brasileiras.....	169

1 INTRODUÇÃO

Redes Geossociais (RGs) são sistemas colaborativos que utilizam a informação geolocalizada como principal elemento na interação. O conteúdo geolocalizado representa a posição real dos usuários na esfera terrestre, encapsulada em mídias como textos e fotos, por exemplo. Esse tipo de dado possibilita uma comunicação entre elementos do mundo digital com elementos mundo real, ou seja, as estruturas geográficas presentes no conteúdo compartilhado correspondem às estruturas existentes no mundo real.

A associação entre esses dois mundos permite o aparecimento de aplicações mais abrangentes como, por exemplo, RGs de controle de emergências naturais em tempo real (KONG, LIU e HUANG, 2014; RÖSLER e LIEBIG, 2013), RGs sobre informações turísticas (CHON, 2013; ZHENG, XIE e MA, 2010), RGs de compartilhamento de rotas para prática de esportes (PERKINS, 2015; SKARBERG, SLETTEN e KOLBEINSVIK, 2014) e RGs sobre sensoriamento participativo (GUO *et al.*, 2016; LI *et al.*, 2016).

Outra característica desses sistemas geolocalizados se refere às novas formas de utilização do conteúdo compartilhado. Para analistas urbanos, as RGs servem como um *proxy* do mundo real sobre o comportamento humano e suas interações sociais em larga escala (LIANG *et al.*, 2013). Para usuários finais, as RGs servem para reduzir o tempo de decisão e aprendizado sobre lugares de uma área (GAO e LIU, 2014). Por fim, donos de estabelecimentos públicos utilizam RGs como aliados comerciais na divulgação e obtenção de *feedbacks* dos visitantes (BAUER e STRAUSS, 2016).

Os principais colaboradores de conteúdo geolocalizado nas RGS são seus próprios usuários. Entretanto, nem todos os usuários registrados são ativos na comunidade. Por isso, esta tese define que o principal valor produzido nas RGs é proveniente de um tipo específico de participante, chamado de criador de conteúdo.

A variedade existente entre os criadores de conteúdo em RGs é a temática central desta pesquisa. Neste trabalho, essa variedade é denominada de diversidade e mensura o quanto os colaboradores são diferentes entre si na comunidade, sob diferentes perspectivas. Pesquisas, como a de Zhu *et al.* (2013), relatam que criadores de conteúdo nos sistemas colaborativos, incluindo RGs, apresentam diversidade entre si. Entender quais características explicam a diversidade dos colaboradores das RGs é um dos principais desafios desta tese.

Na literatura, uma parcela reduzida de trabalhos relata a diversidade dos criadores de

conteúdo em regiões diferentes (GEORGIEV, NOULAS e MASCOLO, 2014; YANG *et al.*, 2016). Nesse sentido, algumas pesquisas consideram a região da interação uma das responsáveis pela multiplicidade de criadores de conteúdo geolocalizado. Assim, a presente tese também considera a região da interação na RG um fator importante nos estudos sobre a diversidade dos criadores de conteúdo. Mais especificamente, esta tese foca em regiões relacionadas às cidades brasileiras.

O Brasil é um contexto promissor a ser considerado nas pesquisas sobre RGs, pois representa um mercado importante no consumo de sistemas colaborativos geolocalizados. Entretanto, há poucas investigações sobre a diversidade dos criadores de conteúdo que interagem com RGs nesse país.

Este capítulo apresenta a introdução desta tese, que propõe uma análise observatória da diversidade dos criadores de conteúdo geolocalizado em RGs usadas nas cidades brasileiras. Nas seções 1.1 e 1.2, o problema de pesquisa e a motivação deste trabalho são descritas. Os objetivos principais e específicos estão listados na seção 1.3 e as contribuições são elencadas na seção 1.4. Finalmente, a seção 1.5 apresenta a estrutura dos demais capítulos desta tese.

1.1. Problema de pesquisa

O trabalho de Yang *et al.* (2016) detecta que criadores de conteúdo são tratados homogeneamente pelas RGs, independentemente de sua origem (turista e residente), objetivo na cidade (residente ou viajante) ou característica demográfica. Nesse sentido, o problema de pesquisa desta tese se refere à quantidade limitada de soluções nas RGs que diferenciem seus contribuidores na comunidade. Como consequência, mecanismos internos às RGs, como métodos de tomada de decisão, recomendação de lugares e mineração de dados, se tornam menos eficientes.

Um agravante do problema de pesquisa diz respeito à escassez de estudos voltados para compreensão da diversidade dos criadores de conteúdo em RGs de cidades brasileiras. O número reduzido de investigações sobre a diversidade inibe a elaboração de melhorias para o tratamento homogêneo das RGs frente aos seus colaboradores. Os trabalhos correlatos mostram que a diversidade dos criadores nas cidades brasileiras é pouco explorada.

Por fim, estudos sobre a diversidade dos criadores de conteúdo em RGs foram amplamente discutidos na literatura, como as investigações de Lee e Chung (2011), Jin *et al.* (2016), Yang *et al.* (2016) e Hristova *et al.* (2016), por exemplo. A falta de padronização na

metodologia e nas métricas desses trabalhos dificulta a investigação sobre a diversidade da presente tese. Em outras palavras, não há um processo metodológico padrão entre os estudos desenvolvidos, nem um consenso sobre as características consideradas na definição da diversidade dos criadores de conteúdo.

1.2. Motivação

O presente trabalho tem como motivação a importância das RGs no processo de tomada de decisão sobre locais. No Brasil, por exemplo, 18% da sua população utiliza RGs para pesquisar sobre locais (GIACOMELE, 2014). Esses sistemas acessam características do comportamento social urbano mais eficientemente do que outros métodos de observação (SILVA *et al.*, 2014). A tomada de decisão é um processo de escolha entre diferentes alternativas, baseado em vários critérios e de acordo com o problema (SHARMA e BAWA, 2018).

Como abordado anteriormente, criadores de conteúdo são os grandes colaboradores das RGs. Entender o comportamento desses participantes é imprescindível para propor melhorias nas RGs relacionadas à recomendação, personalização, filtragem de dados, etc. Portanto, é motivador saber que a análise da diversidade serve de subsídio para projetar qualquer uma dessas melhorias.

Outra motivação diz respeito à escolha de cidades brasileiras como cenário de observação dos criadores de conteúdo geolocalizado. Em 2015, o Brasil contava com 79 milhões de contas ativas nas RGs (GIACOMELE, 2014). Em 2018, esses usuários representarão 41,1% da população brasileira (aproximadamente 85 milhões de usuários) (EMARKETER, 2017). Isso representa uma quantidade expressiva de usuários interagindo com sistemas geolocalizados. Embora o Brasil seja um país pouco comentado nas análises de RGs, claramente demonstra potencial no consumo desse tipo de sistema.

1.3. Objetivos

O objetivo principal desta tese é mapear a diversidade dos criadores de conteúdo em RGs usadas em cidades brasileiras. Para isso, alguns objetivos específicos precisam ser alcançados:

- Compreender como as RGs são usadas em diferentes cidades brasileiras;
- Compreender os processos metodológicos que descrevem a diversidade em RGs;

- Compreender as métricas utilizadas para caracterizar a diversidade nas RGs;
- Definir qual processo metodológico utilizar para o mapeamento dos criadores de conteúdo em cidades brasileiras;
- Definir quais características definem a diversidade dos criadores de conteúdo em cidades brasileiras;
- Mapear a diversidade dos criadores de conteúdo de acordo com suas experiências sobre locais;
- Mapear a diversidade dos criadores de conteúdo de acordo com seus comportamentos colaborativos;
- Mapear a diversidade dos criadores de conteúdo de acordo com seus comportamentos espaciais.

1.4. Contribuições

A principal contribuição desta tese é o mapeamento da diversidade dos criadores de conteúdo em RGs usadas nas cidades brasileiras. Os resultados dessa investigação auxiliam o desenvolvimento de mecanismos de personalização, em RGs usadas no Brasil. Nesse sentido, este trabalho é uma das primeiras iniciativas de mapeamento em larga escala de criadores de conteúdo geolocalizado em cidades brasileiras.

Outra contribuição se refere à elaboração de um processo metodológico nas análises elaboradas. Nos trabalhos encontrados na literatura, é perceptível que a diversidade dos participantes de RGs é definida apenas como uma característica. No presente trabalho, a diversidade é investigada em três diferentes abordagens: de acordo com a experiência, os padrões de comportamento ao contribuir e os padrões espaciais dos contribuidores. Vale ressaltar que cada uma das abordagens segue um processo metodológico particular para o mapeamento da diversidade. Isso é possível devido a um conjunto de pesquisas sobre os métodos mais adequados para cada situação.

Um outro ponto marcante dos estudos foi considerar o Brasil como contexto de observação. Essa escolha elimina a preferência de outros trabalhos sobre colaboradores de RGs em cidades amplamente investigadas. Vale ressaltar também que o Brasil é um contexto nunca antes estudado em larga escala.

Por fim, os resultados desta pesquisa também são úteis para outras áreas, além da computação. Na área de urbanismo, por exemplo, a análise da diversidade nas cidades

brasileiras pode auxiliar a detecção das preferências dos tipos de criadores de conteúdo geolocalizado pelas áreas públicas da cidade.

Na área de turística, a observação dos contribuidores de RGs durante um evento da cidade pode ser útil na obtenção de *feedbacks* sobre lugares públicos e na melhoria desses eventos turísticos em larga-escala. Outro ponto interessante seria na observação dos participantes de RGs durante eventos turísticos em relação ao padrão de movimento na cidade.

Na área social, a detecção da diversidade também serve como indicador de áreas díspares socialmente. Os resultados deste trabalho subsidiam a comparação da participação e das impressões dos colaboradores de RGs em cidades com infraestruturas diferentes.

1.5. Sumário da tese

A presente tese está dividida em três partes. A primeira parte se refere à contextualização da temática das RGs e da investigação de trabalhos na literatura sobre a diversidade dos participantes do sistema geolocalizado. A segunda parte do trabalho diz respeito à execução dos experimentos de mapeamento dos contribuidores em RGs brasileiras em três cenários diferentes. Finalmente, a terceira parte é sobre as conclusões do trabalho, junto com as seções de referências e apêndices.

2 ESTADO DA ARTE

RGs são sistemas usados para construir uma infraestrutura a fim de conectar indivíduos, de acordo com a localização e outras propriedades em comum (BHUIYAN, 2013). Outra explicação aceita por este trabalho define RGs como um tipo de rede social que utiliza dados com a localização do usuário (XU, DING e CHEN, 2015). Redes com essas definições possuem diversas denominações na literatura, por exemplo, redes sociais baseadas em localização, redes sociais móveis, serviços geolocalizados e redes sociais com dados geolocalizados, por exemplo.

Segundo Cho, Myers e Leskovec (2011), a informação geolocalizada abre possibilidade para novas perspectivas do conteúdo compartilhado. Uma marcação geolocalizada possui várias descrições relacionadas aos usuários e às localizações, que são os principais temas da RG e estão estritamente associados entre si (ZHENG, 2012).

A utilidade das RGs para seus participantes tem impacto na aprendizagem e na tomada de decisão sobre lugares, como abordam os trabalhos de Yang *et al.* (2016) e Gao e Liu (2014). No contexto das cidades, RGs podem conter informações úteis em áreas como saúde, entretenimento, controle de emergências, etc. Investigar as aplicabilidades das RGs em uma cidade possibilita entender como é importante o auxílio desse tipo de ferramenta para as áreas urbanas. Baseado nisso, executamos uma busca sistemática sobre pesquisas com RGs usadas em cidades.

2.1. RGs nas cidades

Entre março de 2016 e novembro de 2017 foi executada uma busca sistemática nos repositórios acadêmicos da IEEE¹, ACM² e CAPES³, utilizando uma expressão de busca que envolve diversos termos sinônimos às RGs:

("Mobile Social Network" OR "Geo-Social Network" OR "Location-based Social Network" OR "Spatial social network" OR "Location-aware service" OR "GPS-data-driven social networking service") AND ("city")

Os critérios de inclusão dos trabalhos obtidos são: i) devem ser pesquisas a partir do ano de 2010 e ii) devem ser trabalhos acadêmicos sobre RGs, com alguma utilidade para cidades.

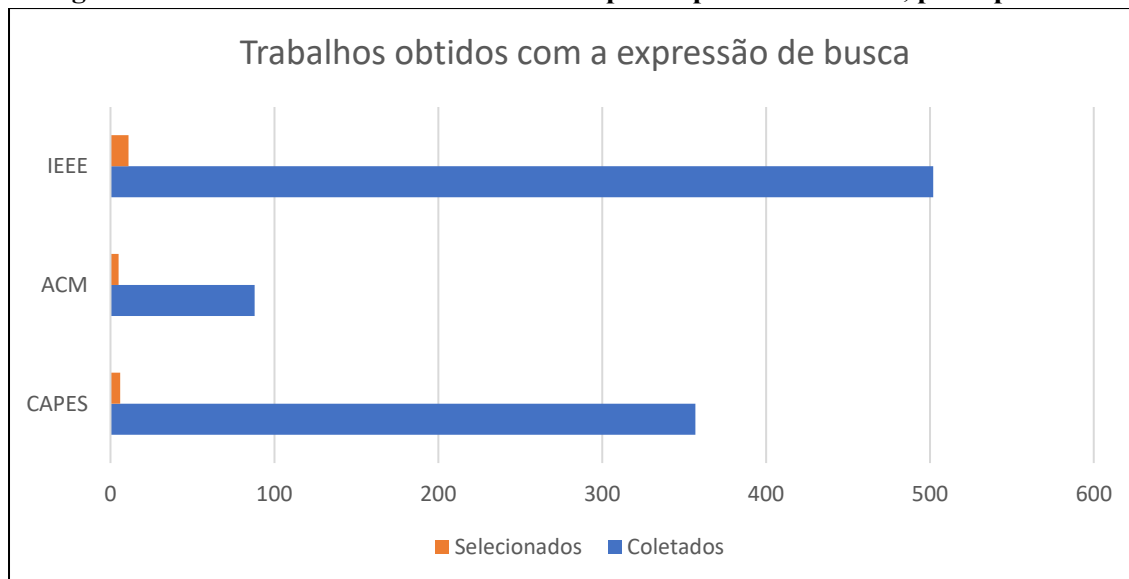
¹ Disponível em: <http://ieeexplore.com>

² Disponível em: <http://acm.dl>

³ Disponível em: <http://capes.gov.br>

Todo o processo de análise das pesquisas é auxiliado pelas ferramentas Evernote⁴ e StArt⁵. Por fim, um resumo sobre os trabalhos obtidos e selecionados está detalhado na Figura 1.

Figura 1. Trabalhos coletados e selecionados pela expressão de busca, por repositório



Fonte: Própria Autora.

Os critérios de inclusão são responsáveis pela descoberta de 33 RGs úteis para cidades, como pode ser visto no comparativo do Quadro 5, no Apêndice A. Embora existam muitos exemplos de RGs comerciais atualmente, esta revisão inclui apenas aquelas discutidas em trabalhos acadêmicos. Portanto, são consideradas RGs úteis para as cidades e que oferecem suporte para áreas urbanas, como turismo, comércio, negócios, saúde, cultura, urbanismo e entretenimento, por exemplo.

A primeira observação se refere às RGs com utilidade na área de urbanismo das cidades. O público-alvo desse tipo de RGs são usuários que utilizam/proveem informações geolocalizadas sobre a cidade. Nesse sentido, se destacam os sistemas PLASH (HO, WU e CHEN, 2010), Askus (KONOMI, 2011) e Socialight (HUMPHREYS e LIAO, 2011). Com o PLASH, a presença de inteligência coletiva para resolver problemas urbanos é uma característica marcante. Em seguida, a RG Askus auxilia participantes com dúvidas sobre características urbanas de uma localidade. No caso do Socialight, há uma troca de experiência entre os habitantes da cidade de Nova Iorque. A interseção entre as RGs do urbanismo é a temática a respeito de soluções urbanas encapsuladas nas informações geolocalizadas.

A próxima RG selecionada é a RG *Lost State College* e representa os sistemas

⁴ Disponível em: <http://evernote.com>

⁵ Disponível em: <http://start.com>

geolocalizados úteis na área cultural da cidade (HAN *et al.*, 2014). Ela objetiva aumentar os laços afetivos dos residentes do Nordeste dos EUA com a história local da cidade, através do compartilhamento de histórias sobre o legado desses locais.

Nos trabalhos selecionados também há RGs ligadas à área de transportes. Elas têm papel ativo no monitoramento e gerenciamento da frota de transportes públicos, por meio de informações sobre horários de pico, detecção de congestionamento, entre outros. Exemplos de RGs nesse enfoque são o TimeTable.Locky (YANO, KAJI e KAWAGUCHI, 2010) e o Moovit (NASH, PURGATHOFER e KAYALI, 2013). Outro exemplo popular nessa área é o Waze, que representa uma iniciativa capaz de prover informações em tempo real sobre o trânsito para o usuário, a partir de colaborações geolocalizadas da comunidade (SILVA *et al.*, 2014).

A saúde é outra área das cidades contemplada pelas RGs. As aplicações utilizadas nesse segmento são úteis para alertar usuários sobre problemas médicos, em áreas com acesso reduzido à saúde, e para informar usuários sobre onde procurar ajuda médica. Alguns exemplos são as RGs iWander (SPOSARO, DANIELSON e TYSON, 2010), *Intelligent Mobile Multimedia Application for the Support of the Elderly* (IMMASE) (PAVLAKIS, ALEPIS e VIRVOU, 2012) e SocialNTT (MUANGNA e NETRAMAI, 2013). Tanto o iWander quanto o IMMASE tentam estabelecer comunicação e monitoramento entre cuidador e paciente. No caso do iWander, os pacientes são pessoas com algum grau de demência e no IMMASE são idosos. Por fim, o SocialNTT auxilia turistas sobre informações médicas na Tailândia.

Na área de entretenimento urbano, alguns exemplos de RGs incentivam usuários a compartilharem localizações para rede de contatos como uma atividade recreativa. Isso acontece em: *Viking Ghost Hunt* (CARRIGY *et al.*, 2010), Instagram (SILVA *et al.*, 2017), Facebook *Places* (HUANG, GALLEGOS e LERMAN, 2017) e Twitter (SOLIMAN *et al.*, 2017). Por fim, elementos de jogabilidade estão presentes na maioria dos exemplos citados.

No turismo, as RGs são úteis no provimento de informações geolocalizadas para usuários com pouca experiência sobre locais da cidade. Isso acontece nas RGs Geolife (ZHENG, XIE e MA, 2010), *Amazing Thailand* (CHON, 2013) e TripPlanner (CHEN *et al.*, 2015). Algumas dessas aplicações focam na troca de experiências entre usuários com preferências (TRAMSNET e SocialTelescope), participações na comunidade (CAMEO) ou trajetórias (Geolife) semelhantes. Outras RGs visam informar sobre locais e acontecimentos turísticos próximos ao local do usuário, como acontece nas RGs *Amazing Thailand*, na Tailândia, e Flarty (DEL BIMBO, FERRACANI e PEZZATINI, 2013), em Florença. Por fim, a RG TripPlanner

contém itinerários de viagens personalizados, de acordo com as preferências dos seus participantes.

O escopo seguinte são RGs para atividades comerciais nas cidades. Algumas delas utilizam informações geolocalizadas para estabelecer relações sociais entre colegas fora da empresa, como a aplicação *Find & Connect* (WANG, ZHU e CHI, 2010). Além disso, outras RGs promovem negócios dentro da comunidade, como bazares virtuais na RG Zaarly (KONOMI, 2011), ou na atração de novos clientes com o PLUTUS (SARWAT *et al.*, 2013).

Os resultados da busca sistemática também mostram que algumas RGs podem contemplar mais de um segmento da cidade. Elas são denominadas RGs polivalentes. Os primeiros exemplos de RGs polivalentes são o TripAdvisor (O'CONNOR, 2010) e o Yelp (BALLESTEROS *et al.*, 2014), pois envolvem setores de turismo e comércio. Esses sistemas disseminam informações sobre qualificação de hotéis e restaurantes, de acordo com avaliações da comunidade, e oferecem mecanismos de monetização, como a aquisição de pontos em programas de fidelidade.

Outras RGs polivalentes privilegiam as áreas de turismo e entretenimento, como SCVNGR (VECCHIONE e MELLINGER, 2011), Broadcastr (SMITH, STRAIGHT e FRANKLIN, 2011), Whrrl (YE *et al.*, 2011) e Sindbad (SARWAT *et al.*, 2012). O sistema mais peculiar entre os selecionados é o Broadcastr porque há uma troca de experiências entre os usuários no formato de áudio. Até o momento, todas as RGs relatadas se apropriam de textos geolocalizados para a descrição dos lugares.

Outra relação existente é entre os segmentos do turismo e da cultura nas cidades. Um exemplo dessa interação acontece na RG *Mobile Guide System Framework for Museums* (SHANG *et al.*, 2011) porque ela permite o aprendizado de visitantes de um museu por meio de uma aplicação que simula um passeio com guia.

Há também uma ligação entre as áreas de saúde e entretenimento. As RGs desse grupo podem ser conhecidas como Jogos *Fitness Móveis (Mobile ExerGames)* (DUTZ *et al.*, 2014), através das aplicações como Strava (NORMAN e KESHA, 2015), Endomondo (SKARBERG, SLETTEN e KOLBEINSVIK, 2014), e Nike *Plus* (PERKINS, 2015). Tais RGs compartilham os hábitos saudáveis dos seus participantes para comunidade, como rotas de corrida, caminhada ou pedaladas, por exemplo.

As últimas RGs polivalentes privilegiam áreas do turismo, comércio e entretenimento. Mais especificamente, as aplicações Foursquare (JAMISON-POWELL *et al.*, 2014) e Gowalla

(SCELLATO e MASCOLO, 2011) adotam recursos com jogabilidade para ensinar turistas sobre locais da cidade e executam promoções e descontos em locais para incentivar os negócios.

Portanto, a análise mostra que RGs podem contribuir diferentemente em vários segmentos das cidades, como mostra o resumo do Quadro 1. As células em cinza representam áreas da cidade que são contempladas por cada RG. Mais detalhes sobre esses sistemas estão presentes Apêndice A.

Quadro 1. Áreas da cidade contempladas pelas RGs

RG	Entretenimento	Comércio	Saúde	Turismo	Transporte	Cultura	Urbanismo
<i>Amazing Thailand</i>							
<i>Askus</i>							
<i>Broadcastr</i>							
<i>CAMEO</i>							
<i>Endomondo</i>							
<i>EKI.LOCKY</i>							
<i>Facebook Places</i>							
<i>Find & Connect</i>							
<i>Flarty</i>							
<i>Foursquare</i>							
<i>GeoLife</i>							
<i>Gowalla</i>							
<i>Instagram</i>							
<i>IMMASE</i>							
<i>IWander</i>							
<i>Lost State College</i>							
<i>Mobile Guide System Framework for Museums</i>							
<i>Moovit</i>							
<i>Nike Plus</i>							
<i>PLASH</i>							
<i>PLUTUS</i>							
<i>Runking</i>							
<i>SCVNGR</i>							
<i>Sindbad</i>							
<i>Socialight</i>							
<i>SocialNTT</i>							
<i>SocialTelescope</i>							
<i>TraMSNET</i>							
<i>TripAdvisor</i>							
<i>TripPlanner</i>							
<i>Twitter</i>							
<i>Viking Ghost Hunt</i>							
<i>Waze</i>							
<i>Whrrl</i>							
<i>YELP</i>							
<i>Zaarly</i>							

Fonte: Própria Autora.

De acordo com uma pesquisa de Arnaboldi, Conti e Delmastro (2011), a atratividade de uma aplicação geolocalizada é identificada por seu impacto nos usuários e na sociedade, criando serviços inteligentes, capazes de ajudar pessoas em suas atividades diárias, melhorar suas interações sociais e oferecer informações úteis customizadas. Pelo impacto visto nas

discussões anteriores dessa seção, é possível afirmar que algumas RGs são suficientemente atrativas para usuários das cidades.

Neste ponto do trabalho, entendemos o impacto positivo que as RGs causam para os participantes das cidades. Mas, existe uma grande variedade de tipos de cidades e, principalmente, de criadores de conteúdo que interagem com esses sistemas colaborativos. Baseado nisso, a próxima etapa do embasamento teórico será entender como a variedade dos criadores de conteúdo nas RGs é abordada nas investigações da literatura.

2.2. Diversidade dos criadores de conteúdo geolocalizado

Diversidade é uma propriedade intrínseca nas plataformas de redes sociais e é dirigida por complexos processos socioeconômicos e técnicos dos sistemas. De acordo com Yang *et al.* (2016), quando a diversidade é negligenciada, há uma dificuldade na generalização e na validade das investigações sobre criadores de conteúdo geolocalizado, causando interpretações incorretas ou ações errôneas por parte dos sistemas colaborativos.

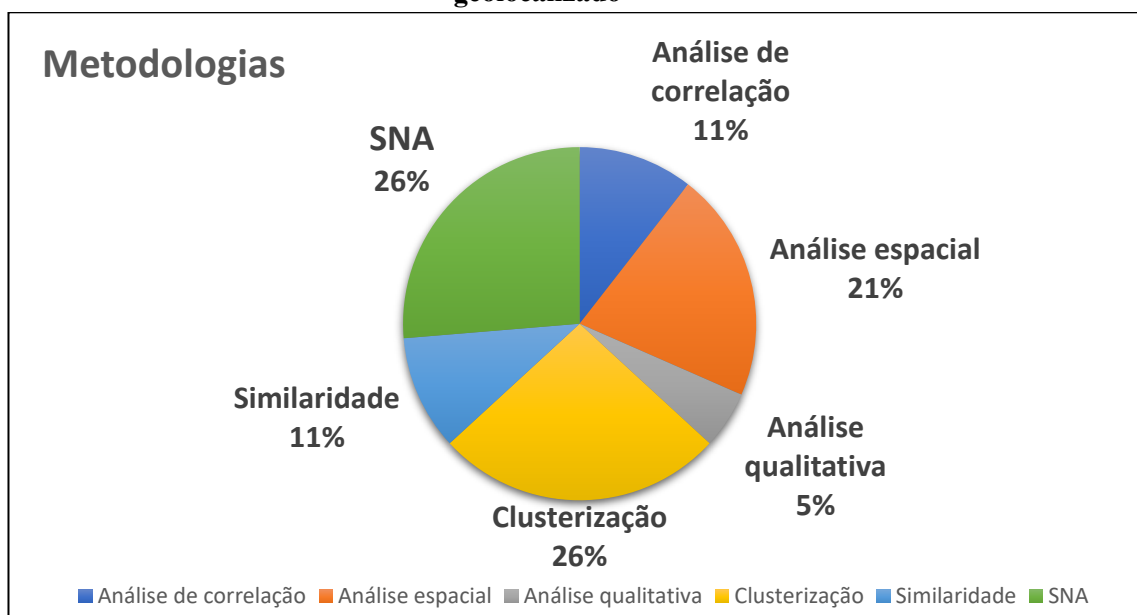
Na literatura, o grande desafio da análise da diversidade dos criadores de conteúdo se resume à escolha de uma metodologia adequada para tal propósito. Essa dificuldade existe porque não há uma padronização no processo, nem nas características utilizadas. Baseado nisso, é investigado como são os processos metodológicos utilizados em pesquisas sobre a diversidade de criadores de conteúdo em RGs.

Para entender os processos metodológicos, foram selecionados 17 estudos entre os anos de 2010 e 2017 sobre a diversidade de participantes em RGs. De acordo com a Figura 2, as metodologias detectadas são: *Social Network Analysis* (SNA), análise de similaridade, clusterização, análise espacial e análise qualitativa.

O primeiro grupo de trabalhos observa a diversidade sob o enfoque das técnicas de SNA. Ela é um processo de investigação das estruturas sociais, através do uso de redes e teoria dos grafos (SCOTT, 2017). Alguns estudos dessa abordagem segmentam usuários por localidade, no que eles chamam de diversidade social do lugar (HRISTOVA *et al.*, 2016). Outros trabalhos definem os participantes das RGs como nós do grafo (LACIC *et al.*, 2014; SILVA *et al.*, 2014). Dessa forma, o SNA representa quão variado um nó e suas conexões podem ser para sua vizinhança. Sarkar, Sieber e Sengupta (2016) apontam que análises de SNA podem ser limitadas em alguns escopos. A crítica sobre essa metodologia é quando ela prioriza os locais onde participantes interagem com RGs, em análises sobre a diversidade. Isso significa que a

validade dos resultados é atingida quando existe uma região de observação com várias localidades incluídas. Dessa forma, amostras aleatórias de lugares não devem usar SNA para mapear a diversidade dos usuários, a menos que considerem o globo terrestre como região de observação, como acontece na investigação de Lee e Chung (2011).

Figura 2. Metodologias presentes nos trabalhos sobre a diversidade dos criadores de conteúdo geolocalizado



Fonte: Própria Autora.

Outro formato metodológico é baseado na análise de similaridade entre participantes das RGs. Essa técnica usa informações dos colaboradores para estabelecer comparativos baseados em métodos como similaridade dos cossenos (GEORGIEV *et al.*, 2014), soma dos pesos (PARK *et al.*, 2017), coeficiente de Jaccard (LACIC *et al.*, 2014), entre outros. A análise de similaridade é útil em estratégias de personalização dos colaboradores da RG, na melhoria das recomendações e na descoberta de comunidades com participantes de interesses semelhantes.

A próxima estratégia metodológica é o processo de clusterização dos criadores de conteúdo geolocalizado. Essa técnica agrupa os participantes de RGs de acordo com um conjunto de variáveis pré-estabelecidas (JIN *et al.*, 2016; YANG *et al.*, 2017). A clusterização é vantajosa para extrair perfis de usuários que são recorrentes na comunidade. Entretanto, ele também tende a desconsiderar participantes que são atípicos, como um super colaborador, por exemplo. Na maioria dos casos, tal comportamento anômalo pode representar *outliers* na amostra analisada. *Outliers* são observações que destoam das demais da amostra, que referenciam o comportamento médio (SEBERT, 1997). Mas, também existe a possibilidade de ser um perfil de usuário relevante e minoritário nas RGs.

A análise espacial também representa uma metodologia de análise da diversidade dos criadores de conteúdo. Muitas das investigações são responsáveis por detectar preferências e padrões relacionados às trajetórias e localidades visitadas pelos contribuidores (ANSELIN e WILLIAMS, 2016; CHORLEY, WHITAKER e ALLEN, 2015; LINS, PEREIRA e BENEVENUTO, 2014; SILVA *et al.*, 2014). O estudo de Chorley, Whitaker e Allen (2015), por exemplo, investiga a diversidade dos usuários sob a ótica dos hábitos individuais, provenientes da associação entre a personalidade do participante e o padrão de movimento que ele percorre com a RG. Outro escopo das análises espaciais detalha a diversidade dos usuários, em relação às trajetórias executadas numa área de observação (ANSELIN e WILLIAMS, 2016; LONG, JIN e JOSHI, 2012).

Outra estratégia é baseada na análise qualitativa das características dos colaboradores para definir a diversidade dos participantes. Esse tipo de metodologia investiga principalmente conteúdos compartilhados na informação geolocalizada (CRANSHAW *et al.*, 2012) e categorias de locais visitados pelos participantes (HUMPHREYS e LIAO, 2011). Assim, a análise qualitativa é eficiente na obtenção de respostas que uma avaliação quantitativa não é capaz de obter.

Por fim, existem trabalhos que utilizam mais de um formato metodológico nas suas investigações sobre a diversidade, como o trabalho de Lee e Chung (2011) que utiliza a semântica dos locais visitados para calcular a similaridade dos participantes das RGs. Outro exemplo é o trabalho de Jin *et al.* (2016) que agrupa os usuários de RGs a partir de métricas baseadas em SNA.

Portanto, a diversidade não é apenas um desafio, mas também uma fonte valorosa de informação, que deveria ser explorada a fim de melhorar os mecanismos interativos nas RGs. Baseado nisso, a próxima fase do trabalho investiga as características que detectam essa heterogeneidade dos criadores de conteúdo geolocalizado, já utilizadas em trabalhos correlatos.

2.3. Métricas da diversidade de criadores de conteúdo geolocalizado

Existem métricas capazes de definir a diversidade dos criadores de conteúdo geolocalizado. Elas estão enquadradas nos seguintes grupos: a) variáveis baseadas nas características dos colaboradores, b) variáveis baseadas no conteúdo compartilhado (HARDY e LINDTNER, 2017), c) variáveis baseadas nos padrões de comportamento ao contribuir (HAWELKA *et al.*, 2014) e d) variáveis baseadas no comportamento espacial (HJORTH, WILKEN e GU, 2012; HONIG e MACDOWALL, 2017).

As variáveis relacionadas às características dos criadores de conteúdo são adotadas em investigações demográficas, como gênero (HRISTOVA *et al.*, 2016), faixa etária e cidade natal (KANG e LERMAN, 2011), por exemplo. Normalmente essas métricas são úteis para iniciar as estatísticas descritivas da amostra.

Os trabalhos que se baseiam no conteúdo compartilhado de colaboradores têm como principais métricas a relevância (BROWN *et al.*, 2014) e a utilidade (JIN *et al.*, 2016) da informação compartilhada na comunidade. Uma pesquisa de Lacic *et al.* (2014) utiliza a relevância dos participantes em três dimensões: social, transacional e localizada. Um criador de conteúdo influente é aquele que compartilha informações geolocalizadas com significado para outros participantes da RG. Nos trabalhos observados, essas métricas são empregadas para mapear preferências e padrões dos contribuidores na comunidade (LACIC *et al.*, 2014; LI *et al.*, 2012).

Outro grupo de trabalho utiliza métricas sobre os padrões de comportamento ao contribuir dos criadores de conteúdo das RGs. A primeira variável desse escopo mensura quantitativamente a colaboração nas RGs e é chamada de atividade ou participação (YANG *et al.*, 2017). Ela representa o número absoluto de interações, em um intervalo de observação, como postagens, *check-ins*, fotos e comentários, por exemplo. Estudos adotam essa variável para obter *feedbacks* sobre a intensidade da colaboração nos lugares (CHANG e SUN, 2011; LI *et al.*, 2013) e nos grupos de usuários (LIANG *et al.*, 2013).

Outra métrica selecionada também se refere à colaboração e se chama variedade. Ela representa a quantidade de locais distintos que um criador de conteúdo visita nas interações com a RG. Estudos adotam a variedade para descrever preferências dos criadores (LI, LI e SHAN, 2017) e para mensurar a distribuição da diversidade pelo índice de Shannon (NOULAS

et al., 2011), importante nos estudos sobre a dinâmica das cidades (PREOȚIUC-PIETRO e COHN, 2013).

Relacionado ao tempo, a variável seguinte avalia o intervalo entre colaborações consecutivas e é conhecida como periodicidade. Tal métrica tem sido utilizada para capturar aspectos da mobilidade humana, usando trajetórias geolocalizadas (RAHIMI e WANG, 2013). É usada para mineração de usuários (RAHIMI e WANG, 2013) e para a definição de parâmetros usados na recomendação (SAKER e EVANS, 2016).

Em relação aos estudos que observam o comportamento espacial dos colaboradores, uma métrica é responsável por mensurar a distância percorrida pelos participantes, durante interações com a RG. Essa métrica é usada no mapeamento de rotas (SALESSES, SCHECHTNER e HIDALGO, 2013) e na identificação dos Pontos de Interesse (POIs) (LINS *et al.*, 2014; SARKAR, SIEBER e SENGUPTA, 2016). Por fim, ela também representa a distância social e contiguidade, em termos geográficos (SCHAAR, VALDEZ e ZIEFLE, 2013) e em termos de relacionamento social (SILVA *et al.*, 2013a; SILVA *et al.*, 2013b).

Outra variável mensura a notoriedade dos locais colaborados para a comunidade e é denominada popularidade. Ela é útil para anunciantes durante a seleção de locais onde aplicar publicidade localizada, e para donos dos locais públicos na obtenção de *feedbacks* sobre a participação da comunidade nos lugares (MOHAMED e ABDELMOTY, 2017; SILVA *et al.*, 2014). Outra linha de trabalho adota a popularidade na construção de uma rede de POIs (WU, MAMOULIS e SHI, 2015). De acordo com Silva *et al.* (2014), algumas características dos locais afetam a popularidade nas RGs, como: a) perfil do lugar, b) categoria do lugar e c) idade do lugar.

Finalmente, a última métrica selecionada diz respeito às avaliações dos locais, calculada pela própria RG. Essa métrica é baseada no número de colaborações, *feedback* dos usuários, entre outras características. Normalmente, a avaliação do local é representada por um índice (WU e LI, 2016). Em análises prévias, métricas de avaliação dos lugares são utilizadas no mapeamento de preferências e padrões dos colaboradores (BROWN *et al.*, 2014).

Todas as métricas abordadas apresentam relevância nas investigações sobre a heterogeneidade dos colaboradores nas RGs. Contudo, a maioria dos estudos selecionados utilizam tais métricas isoladamente. O presente trabalho destaca as investigações que adotam um conjunto de variáveis para mapear criadores de conteúdo das RGs (BROWN *et al.*, 2014; CHEN *et al.*, 2016) porque proveem resultados mais aprofundados sobre a diversidade.

Por fim, os trabalhos investigados apresentam algumas peculiaridades quanto à delimitação espacial. Existem pesquisas que não definem uma área de observação para caracterizar os participantes das RGs (CHORLEY, WHITAKER e ALLEN, 2015; SARKAR, SIEBER e SENGUPTA, 2016; SILVA *et al.*, 2014). Entretanto, há outras investigações que utilizam uma ou mais de uma área de observação nas análises (CHEN *et al.*, 2016; WU, MAMOULIS e SHI, 2015). Nesse último cenário de estudos, a maioria dos experimentos analisam cidades na Europa (CHANG e SUN, 2011; CHEN *et al.*, 2016) e EUA (CHEN *et al.*, 2016; WÖRNDL, HEFELE e HERZOG, 2017; WU, MAMOULIS e SHI, 2015). O restante dos experimentos selecionados é sobre áreas da Ásia (YANG *et al.*, 2016; YANG *et al.*, 2017) e América do Sul (BROWN *et al.*, 2014). Portanto, é perceptível como as investigações sobre criadores de conteúdo geolocalizado são concentradas nas mesmas localidades.

2.4. Discussão do embasamento

A seção de embasamento teórico aborda temas relacionados às RGs para contextualizar a área de atuação do presente trabalho. Primeiramente, é observado quais são as aplicabilidades das RGs nas cidades e, em seguida, quais são as metodologias e métricas utilizadas para mapear a variedade dos participantes nos sistemas colaborativos geolocalizados.

Em relação às RGs nas cidades, é perceptível as inúmeras aplicabilidades que esses sistemas podem trazer para diferentes segmentos urbanos. Com o passar dos anos, houve um aumento no número de pesquisas que utilizam RGs como sensores de fluxo humano. Isso significa que cada vez mais elas têm sido usadas na coleta do comportamento espacial. Entretanto, muitos desses trabalhos são centralizados em áreas já exploradas do globo terrestre e o Brasil não está incluído neste grupo.

Relacionado aos colaboradores das RGs, muitas investigações abordam o comportamento dos criadores de conteúdo geolocalizado. A característica fundamental dos participantes nas RGs é a heterogeneidade entre si na comunidade. Há uma variedade na forma como os colaboradores percebem, experimentam e interagem com a RG. Embora a diversidade dos criadores de conteúdo geolocalizado seja amplamente conhecida, existe pouca padronização nos mecanismos de mapeamento da diversidade e nas métricas utilizadas.

Por tudo que foi discutido, será exposto no próximo capítulo o mapeamento da diversidade dos criadores de conteúdo geolocalizado em cidades brasileiras.

3 DADOS USADOS PARA A ANÁLISE DA DIVERSIDADE DOS CRIADORES DE CONTEÚDO

Baseado no exposto nas seções anteriores, a diversidade dos criadores de conteúdo pode ser mapeada considerando diferentes abordagens. Neste trabalho são exploradas três abordagens para o mapeamento da diversidade. Num primeiro experimento, o mapeamento será de acordo com as experiências dos colaboradores sobre os locais visitados com a RG. Num segundo momento, essa diversidade é pesquisada sob a ótica dos padrões de comportamento ao contribuir dos colaboradores das RGs nas cidades brasileiras. Por fim, a terceira fase representa a análise da diversidade dos criadores de conteúdo em relação aos seus hábitos espaciais com a RG. Detalhes sobre cada fase do mapeamento estarão nos capítulos a seguir e pode ser visto na linha do tempo do Apêndice B .

A escolha dessas abordagens se baseia na análise crítica dos trabalhos correlatos. Inicialmente, métricas relacionadas à experiência e aos padrões de comportamento ao contribuir são selecionadas porque estão presentes na comunidade científica para descrever aspectos da diversidade. Num segundo momento, as discussões dos experimentos com as duas abordagens anteriores definem o surgimento de uma última abordagem: a diversidade baseada no comportamento espacial.

Por se tratar de um estudo exploratório, existem muitos direcionamentos sobre as abordagens. Nesse sentido, nem todas as escolhas são bem-sucedidas no planejamento do mapeamento da diversidade. Inicialmente, a ideia é utilizar a temporalidade das informações geolocalizadas como fator da diversidade. Entretanto, essa possibilidade é descartada porque as amostras coletadas não apresentam dados contínuos de um mesmo criador de conteúdo por um longo tempo. Por isso, a estratégia adotada por este trabalho é a exclusão da temporalidade como fator utilizado no mapeamento.

As análises de mapeamento apresentam alguns pontos de similaridade entre si. Um deles se refere à escolha do cenário de observação dos dados. Em todos os experimentos, a amostra utilizada é composta por usuários que participam da RG em cidades brasileiras. Uma das razões para a escolha do Brasil é que 38% da sua população possui contas em RGs (GIACOMELE,

2017) e isso representa aproximadamente 77 milhões de usuários. Além de esse país ser o contexto de vivência da autora da presente tese, outro fator motivador para a escolha das cidades brasileiras é a quantidade reduzida de estudos sobre sistemas colaborativos nessas localidades, como já mencionado nas discussões da seção 2.4. As cidades brasileiras escolhidas nos experimentos variam de acordo com o propósito de cada observação.

O presente trabalho foca em RGs úteis para o processo de tomada de decisão sobre lugares na cidade. Baseado nisso, todos os experimentos utilizam o mesmo exemplo prototípico de RGs para a coleta dos dados. As RGs selecionadas são o Foursquare e o Swarm. Antes de 2014, esses dois sistemas eram apenas um, chamado de Foursquare, e o usuário tinha a possibilidade de compartilhar dois tipos de conteúdo geolocalizado: *check-ins* e dicas. *Check-in* é um tipo de conteúdo geolocalizado, no qual o usuário informa sua posição geográfica apenas para seus contatos da comunidade. Dica é uma informação geolocalizada, no qual o criador de conteúdo compartilha impressões pessoais sobre os locais visitados para toda a comunidade. Atualmente, a RG que compartilha *check-ins* é o Swarm⁶ e a que compartilha dicas é o Foursquare⁷. Essa subdivisão na RG não trouxe ameaça à validade dos dados utilizados nos experimentos porque os dados mantiveram o mesmo padrão. Atualmente, o Foursquare comporta mais 87 milhões de dicas, em mais de 100 milhões de lugares e o Swarm recebe diariamente mais de 8 milhões de *check-ins*, em média (DMR, 2017).

Outro fator motivador para a utilização das RGs selecionadas é a adesão considerável do público brasileiro. Já houveram temporadas em que o crescimento no número de usuários no Brasil cresceu 800% (EXAME, 2012). Por fim, os experimentos também apresentam similaridade na metodologia utilizada. Em todas as etapas do mapeamento, a obtenção de dados seguiu três processos:

1. Coleta de *check-ins*:

- *Entrada de informações para coleta*: Latitude e longitude das cidades escolhidas e o raio de alcance;
- *Saída dos dados*: arquivos no formato *json* com detalhes sobre os *check-ins*;

⁶ Disponível em: <https://pt.swarmapp.com>

⁷ Disponível em: <https://pt.foursquare.com>

- *Instrumentação*: API do Twitter, pois permite uma coleta de dados com delimitação de localidade. São coletadas as interações da RG que são integradas com o Twitter.

2. Coleta de dicas:

- *Entrada de informações para coleta*: informações sobre os locais para obter o histórico de dicas;
- *Saída dos dados*: arquivos no formato *json* com detalhes sobre as dicas dos lugares solicitados;
- *Instrumentação*: API do Foursquare. Depois de 2014, ficou conhecida como API do Swarm, mas não sofreu alterações na sintaxe.

3. Mineração dos dados:

- *Entrada*: Conjunto de dados de *check-ins* ou de dicas (depende do experimento);
- *Saída*: Tabela mineradas para as análises;
- *Instrumentação*: *Scripts* em Python.

Apesar de similaridades no processo de coleta de dados, o contexto das amostras utilizadas é distinto nos três experimentos. A análise da diversidade de acordo com a experiência utiliza *check-ins* e dicas antes, durante e depois de um evento turístico numa cidade brasileira. A análise da diversidade frente aos padrões de comportamento ao contribuir usa dicas realizadas em diversas cidades brasileiras e num longo período de observação. Por fim, a análise espacial da diversidade se apropria de *check-ins* realizados em diversas cidades brasileiras durante uma longa temporada. Mais detalhes sobre esses dados serão detalhados nos próximos capítulos.

A principal ameaça à validade dos experimentos se refere aos dados selecionados, como acontece em outras investigações que trabalham com *big data* (BOYD e CRAWFORD, 2012; HONIG e MACDOWALL, 2017). O primeiro ponto crítico dos experimentos é sobre a privacidade das informações dos seus usuários. Este trabalho analisa os conteúdos compartilhados e as ações colaborativas dos participantes, sem acessar informações pessoais.

O segundo ponto crítico trata da qualidade do conjunto de dados. Este trabalho mitiga o risco de informações não-confiáveis através da adoção da API do Foursquare, em associação com a API do Twitter. Essa conjunção de APIs é necessária para contornar a limitação de requisições da API da RG na obtenção dos dados.

Discussões estatísticas prévias têm questionado a reprodutibilidade dos estudos científicos, por não apresentarem confiabilidade nos dados (HARLOW, MULAİK e STEIGER, 2016; HIGGINS *et al.*, 2003). De acordo com Benjamin (2017), a principal causa desse problema é que o padrão estatístico das novas descobertas é muito baixo. A associação das descobertas estatísticas com p-valor menor do que 0,05 resulta em falsos positivos. Por isso, é sugerido que experimentos atuais adotem o p-valor menor do que 0,005 para aumentar a reprodutibilidade dos experimentos. Essa sugestão é acatada pela presente tese.

A seguir, será detalhada a primeira abordagem do mapeamento da diversidade, baseada na experiência dos criadores de conteúdo geolocalizado.

4 DIVERSIDADE BASEADA NA EXPERIÊNCIA

Neste experimento, a diversidade dos criadores de conteúdo é relacionada à variedade de experiências dos colaboradores sobre locais da cidade onde interagem com a RG. Nesse sentido, são definidos dois grupos de criadores de conteúdo: turistas e residentes. Turistas são participantes de RGs que colaboram em locais de uma cidade, na qual eles estão temporariamente. Residentes são colaboradores que moram na cidade onde eles interagem com a RG.

Uma ameaça à validade na definição desses grupos é a existência de turistas que compartilham muitas experiências sobre o(s) local(is) visitado(s) e residentes detentores de conteúdos com pouco conhecimento sobre o(s) local(is) de sua cidade de residência. Para que as análises dos grupos gerem resultados consistentes, é importante fazer um controle de *outliers* na amostra. Dessa forma, esses comportamentos “atípicos” nos dois grupos são descartados.

4.1. Relevância do conteúdo geolocalizado

A métrica escolhida para comparar a diversidade colaborativa de residentes e turistas se baseia na aceitação dos outros usuários da comunidade sobre as colaborações realizadas. Essa variável é chamada de relevância do conteúdo geolocalizado.

Um usuário u utiliza a RG Foursquare para compartilhar conteúdos geolocalizados, no formato de dicas. Cada dica c , criada por u , é caracterizada pela tupla $c = \{u, l, h, r\}$, que detalha quem é o autor de c (u), o local onde c foi compartilhada por u (l), o número de colaborações anteriores de outros criadores em l (h) e a relevância do conteúdo de c (r).

A métrica da relevância indica a receptividade de outros participantes sobre o conteúdo compartilhado, fazendo uso de marcações específicas, de acordo com a RG. Baseado nisso, o Foursquare possui marcações – chamados de *done*, *to do* e *like* – que demonstram se a dica compartilhada tem impacto para outros colaboradores da RG. A marcação de *done* denota que o participante é receptivo ao conteúdo da colaboração compartilhada e que ele já esteve no local da dica. A marcação *to do* serve para indicar os usuários que nunca estiveram em um local, mas gostaria de visitar o lugar, motivados pela colaboração geolocalizada. Por fim, *like* denota que um participante é receptivo ao conteúdo compartilhado na colaboração

(LINDQVIST *et al.*, 2011).

A relevância de uma dica c (R_c), de autoria do usuário u , é definida como a razão entre o somatório das marcações de *feedback* da informação compartilhada – no Foursquare são *done*s (e), *to dos* (o) e *likes* (k) – pelo número de dias que aquela informação está exposta na RG (d), como mostra a Equação I.

$$R_c = \frac{e + o + k}{d} \quad (I)$$

A relevância total das informações compartilhadas por usuário u é calculada como a média das relevâncias de colaborações feitas por u e que estão no conjunto de colaborações $C = \{c_1, \dots, c_n\}$, como mostra a Equação II. Baseado no detalhamento da média na Equação III, o usuário u fez n colaborações num dado intervalo de observação.

$$R_u = \mu_{R_C} \quad (II)$$

$$R_u = \frac{\sum_{i=1}^n R_{c_i}}{n} \quad (III)$$

Por fim, a presente tese considera a relevância das dicas normalizadas para explicar a diversidade de turistas e residentes.

4.2. Metodologia

Neste experimento, é importante que turistas e residentes estejam na sua condição de maior equilíbrio quantitativo na região observada. Assim, a observação dos grupos de colaboradores é realizada numa temporada do ano onde existe a maior concentração de turistas na cidade: durante um evento turístico. O evento escolhido para a análise dos grupos é conhecido como São João e está localizado na cidade de Campina Grande. É uma celebração popular que se refere ao nascimento de São João Batista e é uma herança portuguesa desde o período colonial. Esse evento turístico ocorre em todo país, mas é particularmente popular na sua região Nordeste. A cidade de Campina Grande apresenta um evento conhecido como o

Maior São João do Mundo porque celebra 30 dias de festividades no mês de junho. Durante esse mês ocorrem apresentações culturais, shows em áreas públicas e decoração típica por toda a cidade. Uma multidão de aproximadamente dois milhões de visitantes se instala na cidade de 400 mil habitantes durante essa temporada.

Este experimento observa apenas uma cidade porque representa as primeiras delimitações da diversidade da pesquisa e, por isso, é importante ser num contexto conhecido pela pesquisadora.

O comparativo da participação de turistas e residentes, sob enfoque na relevância do conteúdo compartilhado, apresenta alguns questionamentos. São eles:

- Quais fatores afetam a relevância do conteúdo compartilhado pelos grupos de criadores de conteúdo, durante o evento turístico?
- Quão relevantes são as anotações criadas pelos grupos de colaboradores para a comunidade da RG, durante o evento turístico?
- Quão relevantes são as anotações produzidas pelos grupos de colaboradores para a comunidade da RG, depois do evento turístico?

Tais perguntas de pesquisa divide o experimento em três momentos distintos. No primeiro momento, o experimento faz uma análise de correlação sobre quais fatores são responsáveis por tornar dicas de turistas e residentes relevantes na RG. Os fatores mencionados na questão se referem às características da colaboração, como: a) tipo do autor da dica, b) tipo do local visitado pelo criador de conteúdo, c) categoria do conteúdo compartilhado (informação, elogio, crítica, convocação ou dúvida), d) gênero do criador de conteúdo e e) participação prévia do criador de conteúdo, em número de dicas, antes da colaboração. Esses fatores da RG são representados como variáveis *dummies* na amostra, com exceção da participação prévia na RG do autor da dica. Variável *dummy* é uma variável numérica que é proveniente de uma outra categórica (SUITS, 1957).

Para responder o segundo questionamento, é feito um teste de independência das amostras sobre a relevância de dicas de turistas e residentes que utilizam o Foursquare durante o evento turístico. Por fim, é observado se o impacto das contribuições dos grupos na RG se altera após o evento turístico.

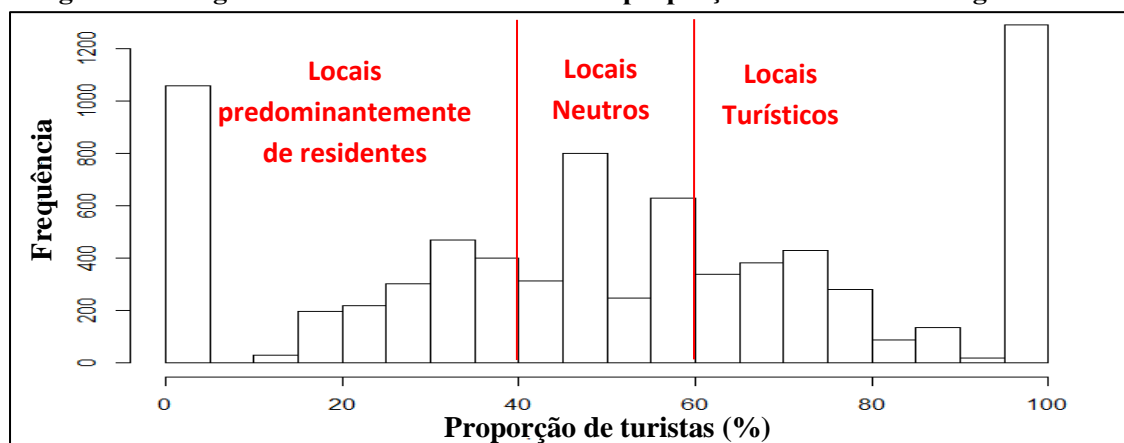
O ponto de partida dessa etapa do mapeamento da diversidade é observar se os grupos de contribuidores lidam com a RG de maneira homogênea. A estratégia inicial é definir que tipo de locais são essencialmente visitados por turistas e por residentes. Mais detalhes na subseção

a seguir.

4.3. Locais de turistas vs. locais de residentes

A amostra das anotações coletadas em cada grupo não detém informações sobre os usuários que utilizam as marcações de *feedback* em cada dica. Assim, a estratégia adotada para saber qual é o público típico de um lugar com a RG, é a observação da predominância de interações na localidade, num dado intervalo de tempo. As interações analisadas são *check-ins* coletados em cada local durante o evento. Baseado nisso, os lugares colaborados pela RG podem assumir uma das três categorias, de acordo com o tipo de participação: a) locais predominantemente turísticos, b) locais neutros e c) locais predominantemente de residentes. A Figura 3 mostra a distribuição dos lugares, baseada na proporção de turistas que visitam as localidades.

Figura 3. Categorias de locais de acordo com a proporção de turistas nos lugares



Fonte: Própria Autora.

Mais detalhes sobre esses tipos de locais, nas subseções seguintes.

4.4. Visão geral dos dados

Os dados observados correspondem a *check-ins* da RG Foursquare realizados entre 7 de junho e 7 de julho de 2013 na cidade de Campina Grande. Os dados obtidos em cada *check-in* são referentes à(ao) data de criação, autor da colaboração, lugar da colaboração e cidade de residência do autor. Um colaborador é considerado turista ou residente de acordo com a informação da sua cidade de residência.

O trabalho de Tung e Ritchie (2011) afirma que conselhos importam e incorporam as

sensações de conforto e desconforto que são relevantes para a revisão de atrações turísticas por parte dos usuários. Por isso, outra informação utilizada nas análises se refere ao histórico de dicas realizadas até agosto de 2013 nos locais marcados pelos *check-ins*. Cada dica contém detalhes sobre o conteúdo compartilhado, o autor da dica e o número de *to dos*, *dones* e *likes*.

A amostra coletada é composta por 2.823 *check-ins* compartilhados em 389 locais diferentes durante o evento, dos quais 1.827 dessas colaborações são criadas por turistas e 996 são pertencentes a residentes. Além do mais, a amostra contém 699 dicas sobre 252 locais distintos, dos quais 123 deles são de lugares turísticos, 70 são em locais predominantemente de residentes e 59 são em lugares neutros. Além do mais, residentes tendem a colaborar, com dicas ou *check-ins*, em locais de alimentação como restaurantes, bares e docerias. Em relação aos locais predominantemente turísticos não há um padrão claro sobre as características dos lugares durante o evento.

A amostra também evidencia que turistas compartilham significativamente mais dicas do que residentes durante o evento turístico. Enquanto isso, residentes compartilham significativamente mais *check-ins* do que turistas, na mesma temporada. Essas constatações são decorrentes do teste comparativo das amostras com o método de Wilcoxon ($p\text{-valor} < 0,005$). Esse método é usado para testar as diferenças entre amostras pareadas de populações (WILCOXON, 1945).

Embora a quantidade de dicas compartilhadas na RG seja diferente entre turistas e residentes, há uma similaridade nos conteúdos compartilhados entre os grupos de colaboradores. De acordo com a Figura 4, uma dica pode conter mais de um tipo de conteúdo, o que justifica a soma das proporções dos conteúdos das colaborações ser excedente a 100%.

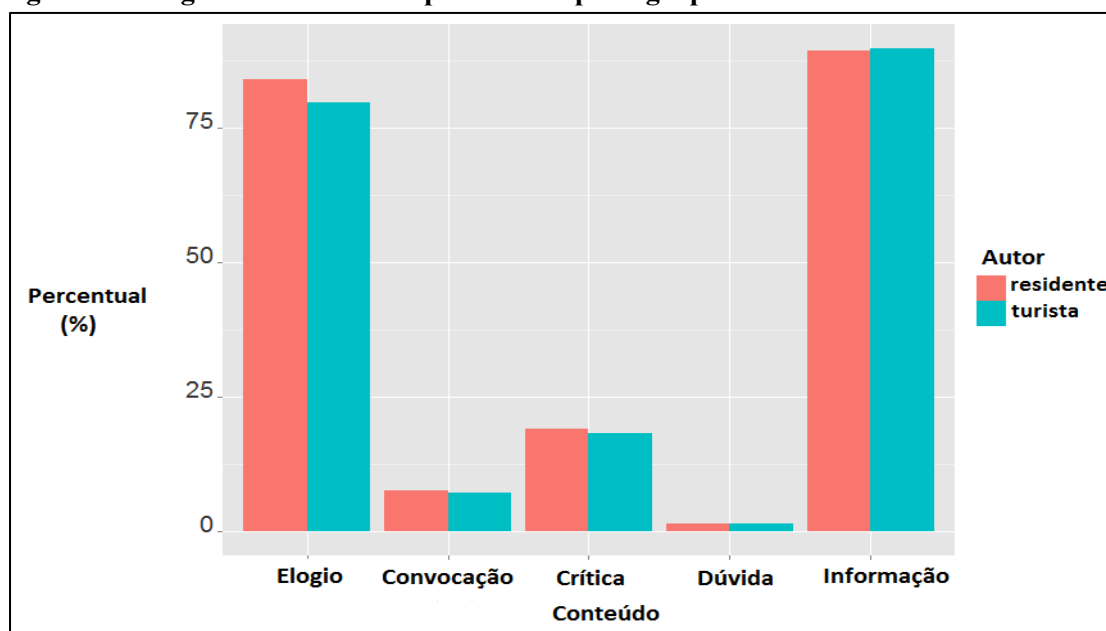
Uma dica apresenta um ou mais tipos de conteúdo encapsulado na informação compartilhada. De acordo com a observação das mensagens geolocalizadas, há sete categorias de dicas. São elas:

- Elogio: expressa satisfação sobre um lugar. Por exemplo: “É um ótimo lugar para se divertir”;
- Crítica: expressa desapontamento sobre um lugar. Por exemplo: “Não venha aqui! A comida é estranha e o preço é caro.”;
- Informação: comunica informações úteis, como senha de rede pública, endereço e horários de funcionamento precisos. Por exemplo: “18:30. Melhor hora para tomar café aqui”;

- Dúvida: contém questionamentos sobre o lugar. Por exemplo: “A taxa de embarque é R\$1,50” ou “eu fui roubada?”;
- Convocação: convida outros usuários a visitarem também um local. Por exemplo: “Vamos provar a melhor comida da região”;
- Ironia: expressa uma mensagem com significado oposto ao escrito. Por exemplo: “a hora do *rush* é a melhor hora para esperar o ônibus aqui. #sqn”;
- Promoção Pessoal: uma mensagem que visa a promoção do próprio autor na comunidade. Por exemplo: “Estou na academia, malhando”.

A definição dessas categorias de dicas se baseia na observação prévia das dicas selecionadas para o experimento.

Figura 4. Categorias de dicas compartilhadas pelos grupos de criadores de conteúdo



Fonte: Própria Autora.

Dois peritos, previamente treinados, são responsáveis por avaliar 100 dicas aleatórias, compartilhadas durante o evento turístico. Um alto grau de similaridade entre as avaliações dos peritos (confiabilidade > 95%) indica confiabilidade nas classificações realizadas nas dicas.

Em seguida, é verificada a relação entre os grupos de colaboradores e as categorias de dicas compartilhadas durante o evento. O método utilizado para essa finalidade é o do teste qui-quadrado de Pearson (SCHEFFE, 1947) e ele avalia quão provável é qualquer diferença observada acontecendo ao acaso. De acordo com os resultados do teste qui-quadrado (p-valor < 0,005), é possível afirmar que algumas categorias de conteúdo podem ser mais recorrentes nas colaborações porque estão associadas ao tipo do grupo do criador. Isso significa que

anotações com promoção pessoal ou ironia são mais prováveis de serem compartilhadas por residentes. Após uma inspeção nos dados, é concluído que dicas com promoção pessoal e ironia apresentam pouca utilidade para o processo de tomada de decisão dos usuários de RG sobre os locais. Como consequência, esse tipo de conteúdo é removido da amostra analisada.

Também são excluídos da amostra os locais particulares, para que turistas e residentes tenham chances iguais de visitar o lugar com a RG. Portanto, das 699 dicas coletadas, restaram 567 dicas sobre locais públicos após a filtragem.

4.5. Quais fatores afetam a relevância do conteúdo compartilhado pelos grupos de criadores de conteúdo, durante o evento turístico?

Essa etapa do experimento investiga os fatores que são responsáveis pela diversidade dos criadores de conteúdo, de acordo com suas experiências e frente à relevância da informação. A estratégia para responder tal pergunta de pesquisa diz respeito à análise da relevância das dicas compartilhadas durante o evento turístico. Para isso, uma análise de regressão é realizada, utilizando os coeficientes normalizados obtidos da relação entre a relevância da dica compartilhada (variável dependente do modelo) e as características da colaboração (variáveis independentes do modelo).

Inicialmente, é verificado se as variáveis independentes exercem influência entre si, em um fenômeno estatístico chamado de multicolinearidade (FARRAR e GLAUBER, 1967), que pode enviesar a análise dos dados. Este trabalho detecta que existe multicolinearidade entre algumas variáveis independentes por meio da execução do teste de correlação não-paramétrico de Kendall (CROUX e DEHON, 2010). Esse método é utilizado para mensurar a associação ordinal entre duas medidas quantitativas.

A correlação entre duas variáveis é existente quando o p-valor é menor do que 0,005 e o índice de correlação (τ) for menor do que -0,30 ou maior do que 0,30. O método utilizado é não-paramétrico porque os dados apresentam comportamento heterocedástico. Retomando o assunto da multicolinearidade, os seguintes pares de variáveis independentes apresentam influência entre si:

- a) Tipo do autor da colaboração e a categoria do lugar colaborado ($\tau = 0,384$ e p-valor $< 0,005$);

- b) Tipo do autor da colaboração e a existência de elogio no conteúdo compartilhado ($\tau = 0,319$ e $p\text{-valor} < 0,005$).

Baseado nos resultados, é descartada a presença de elogio e da categoria do lugar para criação do modelo de regressão. Se essas duas características fossem mantidas na análise, não haveria validade estatística nos resultados.

No que se refere às dicas de turistas, é detectada uma multicolinearidade entre a categoria do lugar compartilhado e a existência de convocação no conteúdo da anotação ($\tau = -0,342$ e $p\text{-valor} < 0,005$). Isso significa que quanto mais turístico um local colaborado for, haverá mais dicas de turistas com convocação encapsulada no conteúdo. Baseado nesses resultados, a existência de convocação é removida da montagem do modelo de regressão.

Em relação às dicas do grupo de residentes, a multicolinearidade está presente entre os seguintes pares de variáveis independentes:

- a) Gênero do autor da colaboração e a participação prévia na RG ($\tau = -0,523$ e $p\text{-valor} < 0,005$);
- b) Gênero do autor da colaboração e existência de crítica encapsulada no conteúdo compartilhado ($\tau = -0,484$ e $p\text{-valor} < 0,005$);
- c) Tipo do lugar colaborado e existência de crítica encapsulada no conteúdo compartilhado ($\tau = -0,553$ e $p\text{-valor} < 0,005$);
- d) Tipos de lugar colaborado e existência de elogio encapsulada no conteúdo compartilhado ($\tau = 0,471$ e $p\text{-valor} < 0,005$).

Baseado nesses resultados, o gênero do autor e a categoria do lugar colaborado são removidos para a elaboração do modelo de regressão, baseado na relevância das dicas. Para a elaboração do modelo de regressão é aplicado o método de regressão binomial negativa (KAPLAN e MEIER, 1958) para mensurar o efeito dos fatores da RG sobre a diversidade dos criadores de conteúdo, de acordo com suas experiências e frente a relevância da informação compartilhada.

A escolha do método de regressão é devido a sua confiabilidade para lidar com variáveis *dummies*, que é a principal característica da amostra analisada. Depois da execução de cada regressão, o coeficiente normalizado de cada fator influente é calculado. São definidos três modelos sobre os fatores que afetam a relevância da informação dos criadores de conteúdo na RG: considerando todas as anotações, apenas anotações criadas por residentes e apenas anotações criadas por turistas. Detalhes sobre esses modelos podem ser vistos na Tabela 1.

Tabela 1. Resultados do modelo de regressão binomial negativa binomial negativa

<i>Grupo de criador de conteúdo</i>	<i>Fator da RG que afeta a relevância da informação geolocalizada</i>	β	<i>p-valor</i>
<i>Residentes + Turistas (n = 148)</i>	O autor é turista? (0 = não e 1 = sim)	- 0,268	< 0,005
<i>Residentes (n = 54)</i>	Colaboração prévia do criador de conteúdo	0,228	< 0,005
	Há convocação no conteúdo? (0 = não e 1 = sim)	- 0,103	
<i>Turistas (n = 94)</i>	Há elogio no conteúdo? (0 = não e 1 = sim)	- 0,197	< 0,005

Fonte: Própria Autora.

De acordo com os resultados, residentes geralmente criam dicas mais relevantes do que turistas durante o evento turístico. Provavelmente, isso ocorre devido ao conhecimento mais aprofundado dos residentes sobre locais de sua cidade de residência. Por exemplo, algumas anotações populares criadas por turistas durante o evento tratam sobre suas experiências iniciais com os lugares:

“O melhor lugar de Campina Grande. Aqui você tem de tudo”.

“Aqui tem a cerveja Antarctica Original, e muito gelada. O feijão com codorna frita é muito bons”.

“As pessoas falam tanto sobre a carne de sol na nata, que eu resolvi provar... é MUITO BOM!”.

“Não venha aqui se você quiser conversar. A música ao vivo é muito alta. Parece mais uma casa de shows”.

Por outro lado, as dicas criadas por residentes durante o evento expressam seu conhecimento mais aprofundado sobre os lugares públicos da cidade:

“A caipirinha vale cada centavo. Se você pedir para fazer forte, se prepare para ficar bêbado”.

“Queijo frito é tradicional aqui. Você pode pedir...”

“Se você está aqui, melhor pedir o galeto completo, com vinagrete e farofa”.

“O creme de camarão em um pão italiano acompanhado de um chá gelado é uma maravilha. Na terça-feira, as pessoas vão lá para jogar xadrez”.

Outra descoberta baseada nos resultados é a influência da participação prévia do autor da dica e da existência de convocação no conteúdo da anotação sobre a relevância do conteúdo

compartilhado pelo grupo de criadores residentes, exclusivamente. Criadores residentes que são participativos tendem a compartilhar dicas mais relevantes do que outros colaboradores que são residentes e colaboram pouco na RG durante o evento turístico. Na direção oposta, anotações de residentes que visam atrair outros colaboradores para um local tendem a ser menos relevantes do que as dicas sem convocação no conteúdo compartilhado.

Outra constatação sobre os resultados do modelo de regressão se refere às anotações com elogios criadas por turistas que tendem a ser menos relevantes do que outras categorias de conteúdo compartilhado. Isso acontece porque o elogio de um turista tem pouco impacto para a tomada de decisão de outros usuários da RG. Por exemplo, há um restaurante com muitas dicas e nesse lugar existe anotações feitas por residentes como: “*O ambiente e o serviço são excelentes. Vale a pena visitar!*” e “*Boa comida, boa variedade e excelente serviço.*”. Nesse mesmo local, existe elogios nas anotações de turistas, como: “*Caipirinha perfeita*” e “*Muito delicioso*”. É possível perceber que há uma precisão maior no conteúdo das dicas elogiosas de residentes do que de turistas.

Quando a amostra contém todos os grupos de criadores de conteúdo, o modelo de regressão detecta que o principal fator da RG responsável por diferenciar a relevância da colaboração dos criadores de conteúdo é o grupo o qual cada colaborador pertence. Pelos resultados, residentes tendem a compartilhar dicas mais relevantes do que turistas. Em locais predominantemente de residentes, a presença de dicas relevantes pode ser ainda mais forte.

Como resposta, os resultados do modelo de regressão apontam a existência de fatores na RG que são responsáveis por determinar a diversidade dos criadores de conteúdo no que se refere às experiências sobre lugares. Quando esses grupos são observados isoladamente, a relevância de suas colaborações é influenciada por fatores distintos. O grupo de criadores residentes participativos tende a criar dicas mais relevantes para a comunidade. Entre os criadores turistas, anotações com elogio parecem ter pouco impacto para a comunidade e os fatores remanescentes não parecem ter significância sobre a relevância da informação compartilhada.

4.6. Quão relevantes são as anotações criadas pelos grupos de colaboradores para a comunidade da RG, durante o evento turístico?

Esta fase da análise tem como objetivo entender se a diversidade dos usuários (de acordo com a experiência sobre lugares) apresenta percepções diferentes sobre o mesmo tipo de conteúdo compartilhado na comunidade da RG. O intuito desse questionamento é avaliar se turistas e residentes consideram como relevantes o mesmo tipo de dica compartilhada durante o evento turístico.

Depois de confirmar que residentes tendem a compartilhar anotações mais relevantes do que turistas, é investigado se isso acontece por causa da audiência nos locais colaborados. Então, a relevância das anotações compartilhadas pelos grupos de criadores de conteúdo é observada em cada uma das categorias de locais, já explicadas na seção 4.3.

Inicialmente, uma análise de multicolinearidade é realizada para considerar os fatores da RG que devem ser incluídos. No que se refere aos fatores da RG relacionados às dicas criadas em locais turísticos, a multicolinearidade estava presente entre a existência de elogio no conteúdo compartilhado e o tipo de autor ($\tau = 0.30$ e p-valor $<0,005$). Isso significa que há uma maior probabilidade de turistas compartilharem dicas com elogio nas anotações de locais turísticos.

Em locais predominantemente frequentados por residentes, a multicolinearidade está presente entre o gênero do autor e a existência de crítica na dica compartilhada ($\tau = -0.43$, p-valor $<0,005$). Isso significa que homens colaboradores tendem a criticar mais nas anotações compartilhadas em locais de residentes. Baseado nesses resultados, o gênero dos criadores de dicas foi removido da análise.

Como na seção anterior, a regressão binomial negativa dá subsídios para responder ao segundo questionamento científico e os resultados desse método estão na Tabela 2. De acordo com o modelo, o fator que tem efeito sobre a relevância diz respeito ao tipo de autor. Notadamente, colaborações de turistas na RG tendem a ser menos relevantes do que colaborações de residentes nas duas categorias de locais (turísticos e de residentes).

Tabela 2. Resultados do modelo de regressão binomial negativa

<i>Categoria do lugar visitado com a RG</i>	<i>Fator da RG</i>	β	<i>p-valor</i>
<i>Local Turístico</i>	O autor é turista? (0 = não e 1 = sim)	- 0,233	< 0,005
<i>Local de residente</i>	O autor é turista? (0 = não e 1 = sim)	- 0,291	< 0,005

Fonte: Própria Autora.

Os resultados apontam que residentes colaboram com dicas mais relevantes para todas as audiências na comunidade da RG. O *feedback* por categorias de lugares representando audiências específicas elimina a relação ser apenas entre amigos da RG.

Por fim, a resposta do segundo questionamento científico é: embora exista uma diversidade de criadores de conteúdo durante o evento turístico, toda a comunidade da RG considera o mesmo tipo de conteúdo como relevante, que são informações compartilhadas por residentes.

4.7. Quão relevantes são as anotações produzidas pelos grupos de colaboradores para a comunidade da RG, depois do evento turístico?

Uma experiência turística pode ser definida como uma avaliação subjetiva individual de eventos que começaram antes, durante ou depois da viagem (TUNG e RITCHIE, 2011). Nesse sentido, o momento de observação final sobre a diversidade dos criadores de conteúdo diz respeito à análise das colaborações dos grupos de colaboradores depois do evento. O objetivo é compreender se as anotações de residentes permanecem mais relevantes do que de turistas, mesmo após o evento turístico.

Para isso, é explorado quais tipos de dicas deixam um legado mais valioso para a comunidade da RG depois do evento. O histórico de dicas é coletado imediatamente após o evento turístico (em julho) e um mês após o evento (em agosto). Esses dois momentos das colaborações são importantes para entender o comportamento das anotações já compartilhadas em instantes diferentes após o evento.

A análise considera dicas sobre locais que permanecem abertos depois do evento. Seguindo esses parâmetros, colaborações de 252 locais são utilizados para o teste não-

paramétrico de Wilcoxon ($p\text{-valor} < 0,005$). Esse método compara as médias das relevâncias das dicas por categoria de local e por coleta pós-evento. De acordo com os resultados do Quadro 2, no geral, as anotações de residentes permanecem mais relevantes do que as dicas de turistas pós-evento, mesmo em locais que são considerados turísticos.

De acordo com comportamento das relevâncias das anotações pós-evento, é percebido que há um decréscimo ao longo do tempo. Assim, o passo seguinte é descobrir se tais relevâncias das dicas compartilhadas no evento turístico diminuem homogeneamente entre os grupos de colaboradores. O teste de Wilcoxon ($p\text{-valor} < 0,005$) é utilizado para comparar essas velocidades de decréscimo da relevância das anotações e os resultados podem ser vistos no Quadro 3.

Quadro 2. Resultados do teste de Wilcoxon

Histórico de dicas do evento em:	Comparação das relevâncias de dicas em:		
	<i>Todos os locais</i>	<i>Locais turísticos</i>	<i>Locais de residentes</i>
Julho	Residentes > Turistas	Residentes > Turistas	Residentes > Turistas
Agosto	Residentes > Turistas	Residentes > Turistas	Residentes > Turistas

Fonte: Própria Autora.

De acordo com os resultados do Quadro 3, as dicas do evento criadas por residentes perdem importância para comunidade de RG mais lentamente do que dicas criadas por turistas. Além do mais, tal comportamento das dicas é percebido em todas as categorias de locais. Os resultados anteriores mostram que residentes compartilham, durante o evento turístico, dicas com validade mais prolongada do que dicas de turistas na comunidade da RG.

Quadro 3. Resultados do teste de Wilcoxon

<i>Locais</i>	<i>Comparações da velocidade de decréscimo da relevância das dicas</i>
<i>Todos os locais</i>	Decréscimo da relevância em dicas de turistas >
	Decréscimo da relevância em dicas de residentes
<i>Locais turísticos</i>	Decréscimo da relevância em dicas de turistas >
	Decréscimo da relevância em dicas de residentes
<i>Locais de residentes</i>	Decréscimo da relevância em dicas de turistas >
	Decréscimo da relevância em dicas de residentes

Fonte: Própria Autora.

O legado das colaborações dos residentes pode ser comprovado no exemplo a seguir, no qual um residente e um turista se referem a um mesmo lugar durante o evento. O turista compartilha o seguinte conteúdo: “Esse bar é muito legal”. De fato, essa dica parece ser útil. Mas, a descrição sobre o lugar não sugere novas experiências. As anotações de residentes expressam suas experiências pessoais com uma aparente confiabilidade: “Tente beber a caipirinha daqui! É a mais barata e a mais deliciosa da região!”. Interpretando o comentário, parece que o criador residente experimentou muitas bebidas e descobriu a melhor dentre elas no lugar.

Outro exemplo é sobre um lugar de lanches muito popular na cidade. As impressões de turistas sobre esse lugar são a respeito da decoração peculiar do lugar, como “Aqui eu estou me sentindo na década de 50” e “A atmosfera daqui é tão antiga”. Por outro lado, as opiniões de residentes sobre o mesmo local se referem às impressões pessoais, como “O bolo de maracujá crocante é delicioso!” e “Lanches e doces são bons, mas o serviço deixa um pouco a desejar”. Assim, os resultados mostram como a diversidade dos usuários, de acordo com a experiência, pode gerar impactos diferentes para a comunidade, mesmo após o período de colaboração.

4.8. Discussão dos resultados

Este experimento executa o mapeamento da primeira abordagem da diversidade dos criadores de conteúdo nas RGs: frente à experiência dos colaboradores sobre os locais visitados. A observação define dois grupos de criadores de anotações para serem analisados: turistas e

residentes. Além do mais, o principal elemento descritor da diversidade dos colaboradores é a relevância da informação compartilhada.

A relevância da informação compartilhada entre os grupos mostra alguns aspectos interessantes da diversidade dos usuários. O principal deles é que as dicas de residentes e de turistas se comportam diferentemente na RG porque são afetadas por elementos distintos da RG. Além disso, o legado das colaborações entre os grupos também é diferente porque residentes compartilham informações com relevância mais duradoura do que as colaborações de turistas.

Entretanto, surgiu alguns pontos de similaridade entre os grupos de criadores de conteúdo. Turistas e residentes consideram dicas de residentes mais relevantes na RG durante o evento turístico. Isso demonstra como turistas aceitam bem as colaborações dos residentes na RG, mas o oposto entre os grupos não acontece.

Portanto, os resultados mostram que a experiência sobre locais é uma característica que pode ser considerada na delimitação da diversidade dos criadores de conteúdo na comunidade da RG.

A próxima etapa será a investigação de uma nova abordagem da diversidade, baseada nos padrões de comportamento ao contribuir dos criadores.

5 DIVERSIDADE BASEADA NOS PADRÕES DE COMPORTAMENTO AO CONTRIBUIR

Esta fase do mapeamento investiga a diversidade dos criadores de conteúdo sob o enfoque dos padrões de comportamento ao contribuir na comunidade. A forma de colaboração dos criadores de conteúdo denota as ações deles na RG Foursquare e em cidades brasileiras. Baseado nisso, alguns questionamentos são levantados:

- Como caracterizar os padrões de comportamento ao contribuir dos criadores de conteúdo de RG?
- Quais são os padrões de comportamento ao contribuir entre os criadores de conteúdo na RG utilizada em cidades brasileiras?

Uma possível ameaça à validade deste experimento se refere à escolha das variáveis que caracterizam a colaboração dos criadores de conteúdo na RG. Por isso, este experimento adota apenas métricas que já foram exploradas por trabalhos na literatura para caracterizar a participação dos criadores de conteúdo da RG. Mais detalhes sobre as variáveis que caracterizam os padrões de comportamento ao contribuir, a seguir.

5.1. Aspectos metodológicos

Um estudo exploratório sobre os padrões da colaboração é realizado entre os criadores de conteúdo na RG Foursquare, utilizadas cidades brasileiras. A clusterização particionada é estratégia utilizada para identificar criadores de conteúdo com comportamentos similares (O'CONNOR e HERLOCKER, 1999). Os algoritmos de clusterização particionada decompõem a amostra dentro de grupos com comportamentos colaborativos distintos entre eles. Dessa maneira, o método de clusterização escolhido é o *k-means* por causa de sua escalabilidade e versatilidade. Esse método particiona as amostras em grupos, de acordo com as médias de referência de cada grupo (JAIN, 2010).

A amostra é composta por dicas compartilhadas em locais de 21 cidades brasileiras (Belém, Belo Horizonte, Brasília, Campinas, Campo Grande, Curitiba, Duque de Caxias, Fortaleza, Guarulhos, Maceió, Manaus, Nova Iguaçu, Recife, Rio de Janeiro, Salvador, São

Bernardo do Campo, São Gonçalo, São Luís, São Paulo, Porto Alegre e Teresina), entre 2013 até 2017.

O intuito deste experimento é observar os padrões de comportamento ao contribuir dos colaboradores num maior número possível de cidades brasileiras. O critério de seleção das cidades se baseia na quantidade mais expressiva de dados, no formato de dicas, coletada por esta pesquisa.

5.2. Padrões de comportamento ao contribuir dos criadores de conteúdo

A análise de clusterização precisa de um conjunto de métricas que descrevam a colaboração dos participantes na RG. Portanto, tais dimensões são baseadas em usos prévios na literatura, que foram adaptadas para o contexto do presente experimento.

Duas métricas selecionadas caracterizam as preferências dos colaboradores pelos lugares: a) a popularidade do lugar e b) a avaliação média do lugar. A popularidade de um lugar representa a atividade no lugar, num dado intervalo de observação na RG. No Foursquare, a popularidade de um local é medida pelo número dicas compartilhadas lá, num dado período de observação. As dicas foram escolhidas para caracterizar a popularidade porque RGs mostram dicas prévias de um lugar antes do criador de conteúdo fazer sua colaboração.

Um criador u compartilha colaborações, no formato de dicas, em um conjunto de locais L , onde $L = \{l_1, l_2, \dots, l_n\}$. Cada local da RG tem sua popularidade P_{li} . Portanto, a preferência por popularidade de u (PP_u) representa a popularidade média dos n lugares visitados por ele, como mostra a Equação IV.

$$PP_u = \frac{\sum_{i=1}^n P_{li}}{n} \quad (IV)$$

A outra métrica mensura as preferências de criadores de conteúdo por lugares, frente à avaliação média do local. Na RG Foursquare, cada local apresenta uma nota entre 0 e 10, calculada pela RG e baseada na informação compartilhada. A nota é gerada de acordo com *feedbacks* implícitos e explícitos dos criadores de conteúdo no local observado.

O criador de conteúdo u visita um conjunto de locais L , onde $L = \{l_1, l_2, \dots, l_n\}$, para

compartilhar dicas. Cada local tem uma nota S_l . Assim, a avaliação média dos locais visitados por u (A_u) é calculada como a média aritmética das notas de todos os n locais visitados por u , como mostra a Equação V.

$$A_u = \frac{\sum_{i=1}^n S_{li}}{n} \quad (V)$$

Outras métricas são escolhidas para caracterizar a participação ao longo do tempo e espaço. A primeira delas é a atividade e representa o número de dicas produzidas pelo colaborador u , num intervalo de observação.

A segunda métrica é a variedade das colaborações, que representa o número de locais distintos que o colaborador visitou para compartilhar dicas com o Foursquare. Um criador u compartilhou dicas em um conjunto de locais L , onde $L = \{l_1, l_2, \dots, l_n\}$. A variedade das colaborações de u diz respeito à quantidade de locais em L que não se repete.

A próxima característica da participação colaborativa tem a ver com a distância percorrida nas colaborações executadas. Um criador u compartilha dicas em um conjunto de locais L , onde $L = \{l_1, l_2, \dots, l_n\}$. Cada lugar visitado tem uma localização geográfica composta por sua latitude e longitude e a distância entre dois pontos, que é calculada pela função $z(l_1, l_2)$, expressa em quilômetros. A observação dos pontos de u , válidos na superfície terrestre, possibilita a definição de um ponto central m com uma posição geográfica também existente. Assim, a métrica da distância entre as colaborações de u (W_u) é calculada como a média das distâncias entre os lugares anotados por ele e o centróide m , como mostra a Equação VI.

$$W_u = \frac{\sum_{i=1}^n z(l_i, m)}{n} \quad (VI)$$

Em seguida, surge uma métrica chamada de periodicidade (R_u) para expressar temporalidade das colaborações do criador. Um colaborador u compartilha dicas num conjunto de locais L , onde $L = \{l_1, l_2, \dots, l_n\}$. Tais dicas foram criadas em instantes de tempo diferentes, t_1, t_2, \dots, t_n . A periodicidade é representada pela média de dias que separam as n colaborações consecutivas de um mesmo criador de conteúdo, como mostra a Equação VII.

$$R_u = \frac{\sum_{i=1}^n (t_{i+1} - t_i)}{n} \quad (\text{VII})$$

A última métrica é a influência social (I_u) de um criador de conteúdo u . Na RG Foursquare, uma dica tem três marcações que expressam a importância da colaboração compartilhada. São elas: a) *like(k)*: uma marcação para dar um *feedback* positivo, b) *to do(o)*: serve para marcar o local que o leitor da dica gostaria de visitar, motivado pelo conteúdo compartilhado e, finalmente, c) *list(y)*: uma marcação que adiciona o local relacionado à dica numa lista pessoal e privada. Embora essas marcações tenham propostas diferentes entre si, elas são igualmente importantes para mensurar a visibilidade e impacto das colaborações realizadas pelos criadores de conteúdo na RG.

Portanto, a influência social de um criador de conteúdo é definida como a média entre o somatório de *likes*, *to dos* e *lists* pelo número de seguidores, como mostra a Equação VIII. O número de seguidores f_u representa o montante de usuários que visualizam a dica do criador de conteúdo e n representa o número de dicas criadas por u .

$$I_u = \frac{\sum_{l=1}^n \frac{k_l + o_l + y_l}{f_u}}{n} \quad (\text{VIII})$$

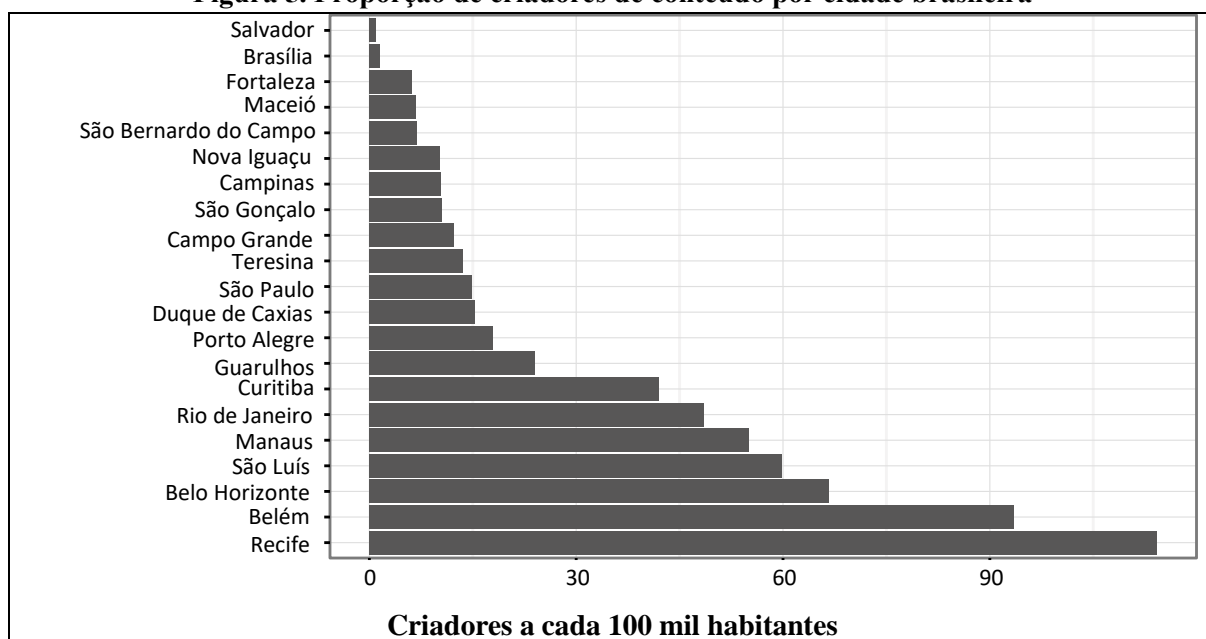
5.3. Visão geral dos dados

A amostra contém 13.783 criadores de conteúdo do Foursquare, responsáveis pela criação de 266.678 dicas em 21 cidades brasileiras. Esses colaboradores criaram, no mínimo, três dicas entre 24 de novembro de 2009 e 22 de agosto de 2016.

A amostra inclui criadores de conteúdo com uma participação até dez vezes maior ou menor que o comportamento médio. Os colaboradores que estão fora dessa margem são considerados *outliers* e extraídos da amostra.

Naturalmente, a incidência de criadores de conteúdo varia de uma cidade brasileira para outra. Esse panorama pode ser visto na Figura 5, que expressa a proporção de colaboradores a cada cem mil habitantes da cidade. A proporção de criadores de conteúdo é a maior na cidade de Recife (114,26 criadores a cada 100 mil habitantes) e menor na cidade de Salvador (0,93 criadores a cada 100 mil habitantes).

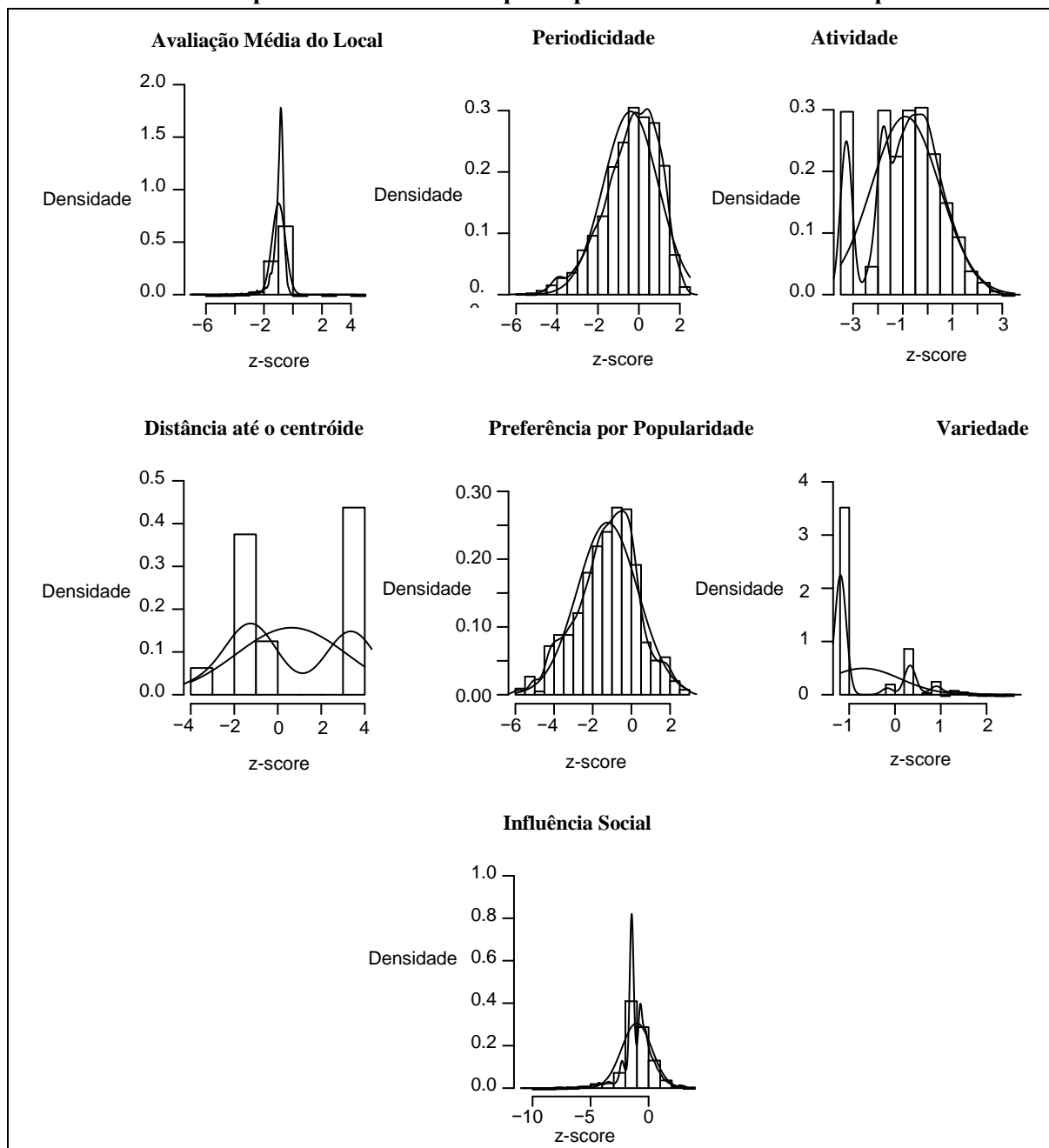
Figura 5. Proporção de criadores de conteúdo por cidade brasileira



Fonte: Própria Autora.

Em média, os criadores de conteúdo compartilharam 6,6 dicas em 1,7 lugares. Outra constatação é que cada colaborador produziu dicas consecutivas com um intervalo de 45,2 dias, a 1,5 km de distância do ponto central de suas colaborações totais. Além disso, os criadores de conteúdo observados preferiram locais com uma participação de 53,9 dicas prévias e com uma nota de 5,1 para compartilhar dicas. A distribuição de todas as métricas que caracterizam os padrões de comportamento ao contribuir pode ser vista na Figura 6.

Figura 6. Histograma das métricas da RG sobre criadores de conteúdo. Todas as métricas estão normalizadas pela técnica do z-score para apresentarem tamanhos comparáveis

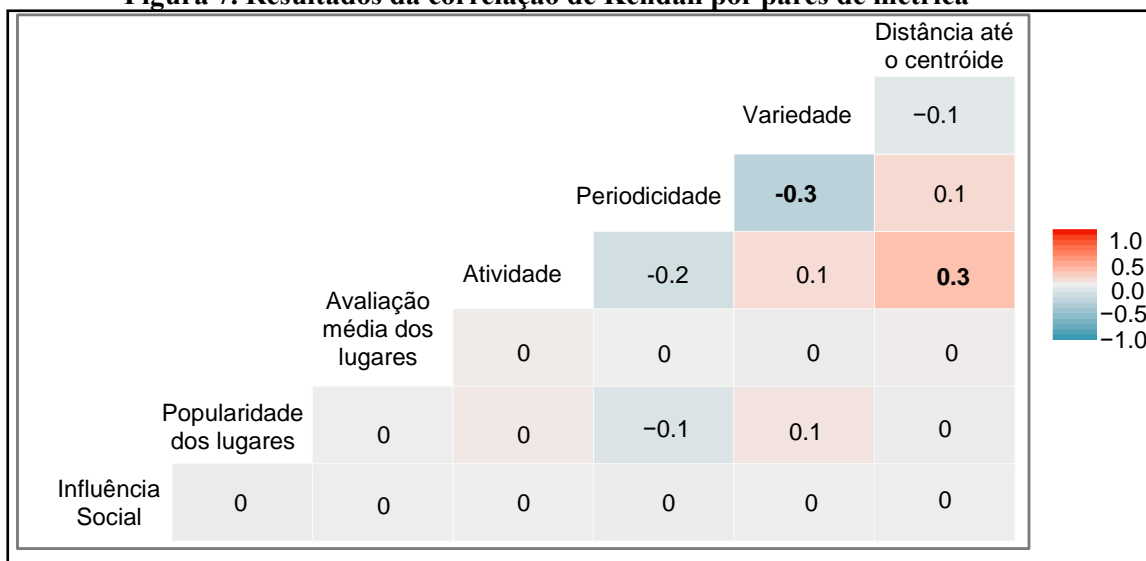


Fonte: Própria Autora.

O passo seguinte é a análise de correlação para explorar a variação de paridade entre as métricas do comportamento participativo. A correlação de Kendall é escolhida por ser um método não-paramétrico que mensura o grau de dependência das variáveis. De acordo com a Figura 7, a maioria das métricas utilizadas não exercem influência entre si, exceto os pares: a) periodicidade e variedade (coeficiente de correlação = -0,3) e b) distância até o centróide e

atividade (coeficiente de correlação = 0,3).

Figura 7. Resultados da correlação de Kendall por pares de métrica



Fonte: Própria Autora.

A primeira relação mostra que, em geral, criadores de conteúdo que visitam mais locais distintos dos outros colaboradores tendem a compartilhar dicas consecutivas num tempo menor do que outros colaboradores. A segunda correlação indica que criadores que compartilham mais dicas tendem a produzir dicas em locais mais distantes.

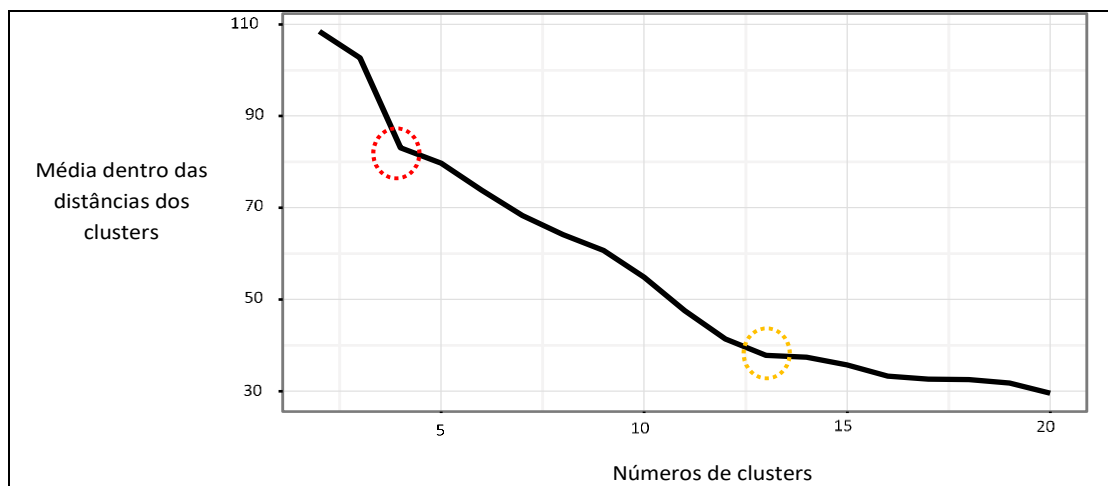
5.4. Definição da quantidade de grupos de colaboradores para clusterização

Na clusterização particionada, é preciso decidir o número de grupos da amostra observada. Há várias heurísticas de decisão sobre o número de grupos mais apropriado. Este experimento adota a técnica do cotovelo para a definição da quantidade de grupos de colaboradores para análise (BHOLOWALIA e KUMAR, 2014). Tal método utiliza a visualização da soma dos quadrados das distâncias entre pontos no grupo e seus centróides, como mostra a Figura 8. O número mais apropriado de grupos é escolhido quando as somas das distâncias começam a decrescer, formando linhas no formato de cotovelo. Essa região do cotovelo representa o número de *clusters* mais adequado para a amostra.

Baseado nos resultados da Figura 8, a melhor configuração do gráfico utiliza quatro *clusters* (círculo vermelho da Figura 8). O gráfico também mostra que uma segunda opção seria de treze grupos (círculo amarelo da Figura 8), mas essa classificação é descartada porque

a análise seria direcionada para a explicação de grupos muito específicos.

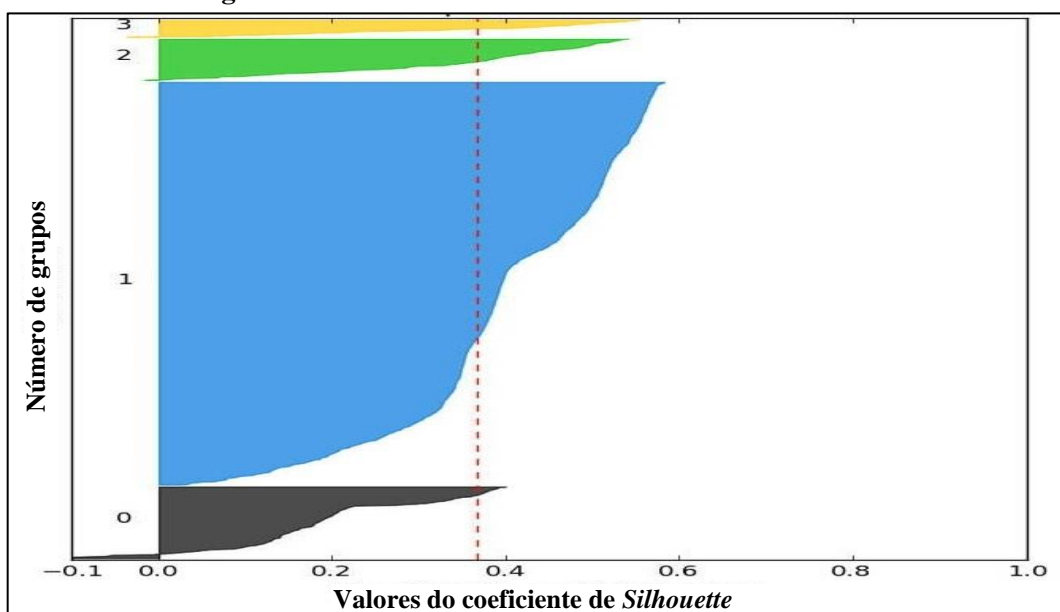
Figura 8. Gráfico sobre a técnica do cotovelo que representa o número de *clusters* pelas distâncias de *cluster* média



Fonte: Própria Autora.

Em seguida, este experimento valida a definição dos grupos com a análise de *Silhouette*, como mostra a Figura 9. Esse método avalia a similaridade de um elemento no cluster, em comparação com outros *clusters* resultantes (ROSSEEUW, 1987). O coeficiente de *Silhouette* mensura quão distante cada ponto é do ponto mais próximo do *cluster* que ele não pertence. O coeficiente *Silhouette* igual a -1 significa que um ponto é muito mais próximo de um grupo o qual ele não faz parte. Analogamente, o coeficiente de *Silhouette* igual a 1 detona que o ponto é muito mais próximo do seu grupo do que dos outros grupos.

Figura 9. Análise de acordo com o método de *Silhouette*



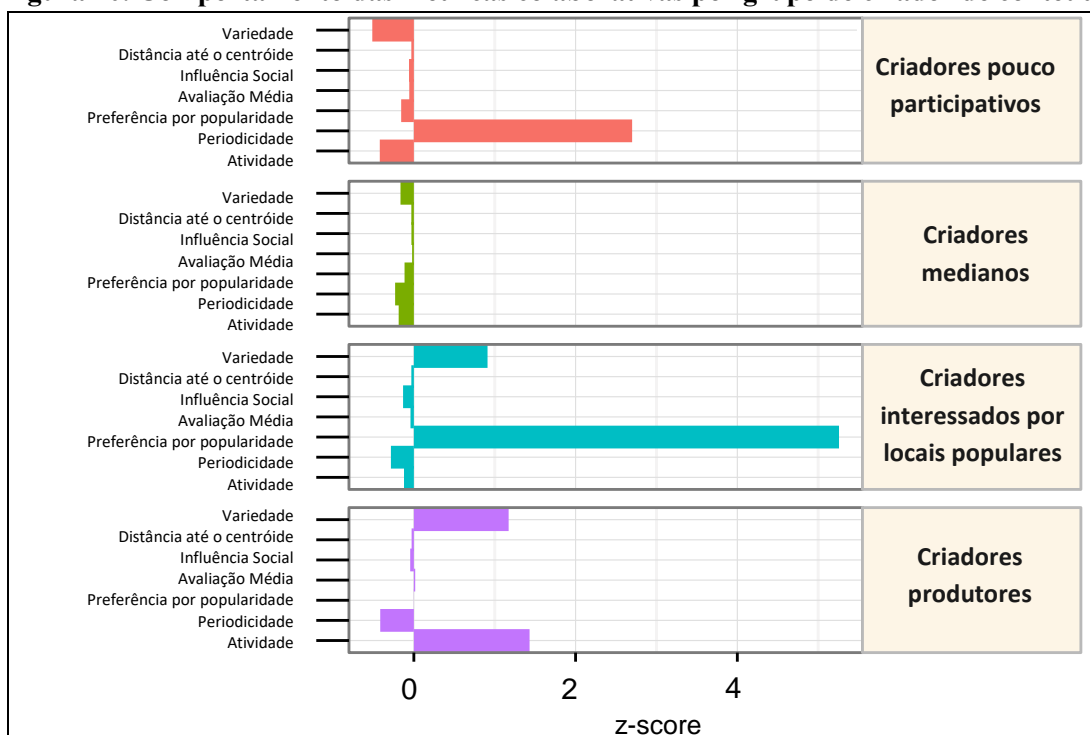
Fonte: Própria Autora.

De acordo com os resultados, a formação de quatro grupos de criadores de conteúdo apresenta índices aceitáveis do coeficiente de *Silhouette*.

5.5. Padrões da participação colaborativa dos criadores de conteúdo de RG em cidades brasileiras

Seguindo a sugestão dos métodos anteriores, o processo de clusterização particionada da participação colaborativa de criadores de conteúdo apresenta quatro grupos. Cada um deles é intitulado de acordo com as características mais aparentes do centróide. Essas nomenclaturas estão ilustradas na Figura 10: criadores pouco participativos, criadores medianos, criadores interessados por locais populares e criadores produtores. É utilizado o método do *z-score* para padronizar as escalas e facilitar a comparação da métricas no gráfico. Por exemplo, o índice 0 implica que os valores estão na média daquela métrica e valores positivos são valores acima da média.

Figura 10. Comportamento das métricas colaborativas por grupo de criador de conteúdo



Fonte: Própria Autora.

A seguir, é detalhado cada perfil de criador de conteúdo geolocalizado, de acordo com as características principais.

5.5.1. Criadores de conteúdo pouco participativos

O processo de clusterização detectou 1.185 criadores de conteúdo pouco participativos, presentes em todas as 21 cidades brasileiras. Esse tipo de colaborador é responsável pela elaboração de 7.306 dicas e tem como principal característica ser o perfil menos participativo na RG, com uma média de 3,43 dicas produzidas. Além do mais, eles demoram o maior tempo para compartilhar dicas consecutivas, com um intervalo de tempo médio de 311,5 dias. Sendo os menos ativos na comunidade, os criadores pouco participativos tendem a visitar menos locais diferentes (média de 1,26 lugares distintos por colaborador).

5.5.2. Criadores de conteúdo mediano

Esse grupo contém 10.173 criadores de conteúdo e representam a maioria dos criadores de conteúdo da amostra. Eles são os colaboradores que oferecem a maior quantidade de contribuições na RG, com 132.313 dicas compartilhadas. De acordo com o centróide desse grupo, criadores medianos compartilham em média 5,66 dicas, um valor próximo ao comportamento geral. Além do mais, eles variam pouco de locais em suas inúmeras colaborações (média de 1,45 locais diferentes por colaborador). É possível dizer que criadores medianos são um público cativo nos lugares que lhes agradam.

Esse perfil de colaborador percorre distâncias próximas ao comportamento mediano (1,53 quilômetros) para criar dicas. Essas colaborações são feitas consecutivamente em um intervalo de tempo menor que um mês (23,80 dias), que é um pouco menor do que a média. No que se refere às preferências espaciais, os criadores de conteúdo medianos escolhem locais com uma popularidade média de 42 dicas por lugar e uma avaliação média de 5,1 de nota, um pouco abaixo do comportamento geral. Isso significa que criadores medianos focam em locais avaliados e frequentados sem grande expressividade, positiva ou negativa, para compartilhar informação geolocalizada. Por fim, criadores medianos têm uma influência social abaixo da média (1,02).

5.5.3. Criadores interessados por locais populares

Esse grupo representa 1.829 colaboradores que tem uma preferência marcante por locais que são bem visitados pela comunidade de RG. Eles estão presentes em 15 cidades brasileiras e são autores de 23.647 dicas da amostra observada. Criadores interessados por lugares populares tende a escolher locais com uma nota de avaliação abaixo da média (4.89). Isso

significa que eles não se baseiam em avaliações providas pela RG para colaborar, pois a popularidade é um critério mais importante nas escolhas deles sobre locais. Talvez, criadores de conteúdo interessados por locais populares prefiram esse tipo de lugar provavelmente porque:

- a) Eles imitam seus amigos: por um princípio conhecido como homofilia (MCPHERSON, SMITH-LOVIN e COOK, 2001), os participantes de uma mesma comunidade adquirem hábitos semelhantes;
- b) Eles apresentam pouco conhecimento sobre as áreas visitadas: eles podem ser colaboradores que estão fora de sua área de costume e apenas seguem o fluxo de atividade da comunidade.

A busca por locais populares faz com que esses colaboradores visitem a maior quantidade de locais distintos de todos os grupos (3,35 locais diferentes). Outra característica é que criadores interessados por locais populares apresentam a menor influência social entre os grupos (média de 0,77). Isso acontece porque a criação de dicas que sejam relevantes em locais populares é uma tarefa mais difícil do que em outros tipos de locais.

Eles também produzem uma soma de dicas abaixo da média (5,42 dicas por criador) e compartilham dicas num intervalo de tempo relativamente rápido (3,87 dias). Em outras palavras, criadores de conteúdo interessados por locais populares produzem menos dicas, mas sua colaboração consecutiva ocorre num intervalo de tempo menor.

Por fim, a participação desses criadores se concentra nas menores áreas porque eles percorrem as menores distâncias de todos os grupos (0,94 km).

5.5.4. Criadores produtores

Os 596 produtores de conteúdo das amostras compartilharam 30.676 dicas em 16 cidades brasileiras, exceto nas cidades de Brasília, Campinas, Salvador, São Bernardo do Campo e Teresina. Eles são os colaboradores mais produtivos, com uma média de 32,09 dicas por criador de conteúdo. Os produtores colaboram numa maior variedade de locais (média de 1,85 lugares diferentes por criador) e gastam o menor tempo de todos os grupos de colaboradores para criar dicas consecutivamente (média de 3,71 dias por criador).

Toda a participação é concentrada em áreas pequenas porque a distância média entre as colaborações de um produtor é abaixo da média (3,10 km). Além disso, produtores preferem locais com popularidade e avaliação medianas (51,07 dicas por local e 4,85 de nota) para

compartilhar dicas. Por fim, esses usuários representam o perfil de colaborador com a maior influência social de todos os grupos e essa boa aceitação por parte da comunidade de RG incentiva a sua alta colaboração na comunidade.

5.6. Impressões sobre os grupos de criadores de conteúdo

Relacionado às RGs, outras características dão *feedback* sobre a visibilidade das contribuições, como as marcações “seguindo” e “seguidores”, pertencentes a cada colaborador. A marcação “seguindo” representa o montante de criadores de dicas que um usuário visualiza as contribuições compartilhadas na RG, no formato de *check-ins* ou dicas.

É descoberto que criadores medianos seguem a maior número de colaboradores na RG (média de 192,61 colaboradores) e os criadores produtores seguem a menor quantidade (média de 163,99 colaboradores). Aparentemente, produtores aprendem menos do que criadores medianos com as informações compartilhadas na RG sobre lugares. Por fim, criadores pouco participativos e criadores interessados por locais populares seguem uma quantidade próxima de colaboradores no Foursquare (média de 176,46 e 177,35; respectivamente).

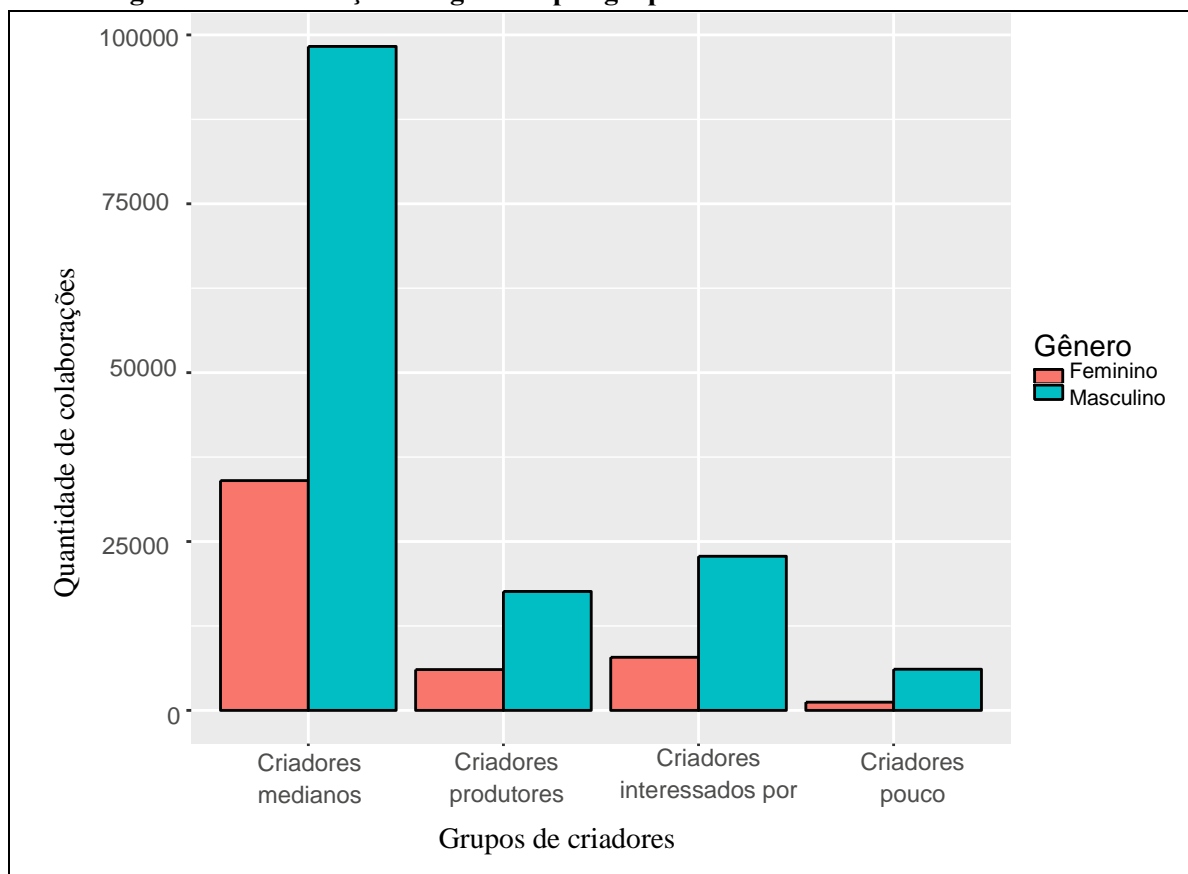
Em relação ao número de seguidores por grupo, produtores têm o maior número de usuários observando suas colaborações (média de 308,17 seguidores) e criadores pouco participativos apresentam um comportamento oposto (média de 185,94 seguidores). Produtores não parecem procurar por opiniões alheias sobre os lugares na comunidade, mas suas impressões pessoais parecem ser bem aceitas pela comunidade da RG. Em relação aos criadores pouco participativos, a falta de interesse deles em compartilhar poucas dicas pode ser motivada pela pouca visibilidade das suas contribuições na comunidade. Por fim, criadores medianos têm um número de seguidores similar aos criadores interessados por locais populares (média de 249,19 e 236,59; respectivamente).

Outro caminho deste mapeamento se refere à observação do tipo de conteúdo compartilhado nos perfis de criadores. Com a utilização do método TF-IDF (*Term Frequency – Inverse Document Frequency*) para atribuição de pesos aos termos mais importantes nos conteúdos geolocalizados, é verificado que cada perfil de criador tem particularidades no conteúdo compartilhado. Por exemplo, criadores interessados por locais populares compartilham informações em que a palavra “pessoas” é o termo mais importante. Isso comprova o interesse desses colaboradores por locais povoados. Entretanto, algumas similaridades entre os grupos também são existentes. Por exemplo, as dicas dos quatro perfis

consideram as palavras “lugar”, “atendimento” e “comida” como termos importantes. Isso indica que criadores de conteúdo nas RGs de cidades brasileiras sentem-se confortável em compartilhar dicas sobre comida, bem como aspectos sobre o lugar.

Naturalmente, todos os grupos apresentam colaboradores masculinos e femininos, como mostra a Figura 11. Mas, criadores homens são maioria em todos os grupos e a maior diferença quantitativa entre os gêneros ocorre entre os criadores medianos.

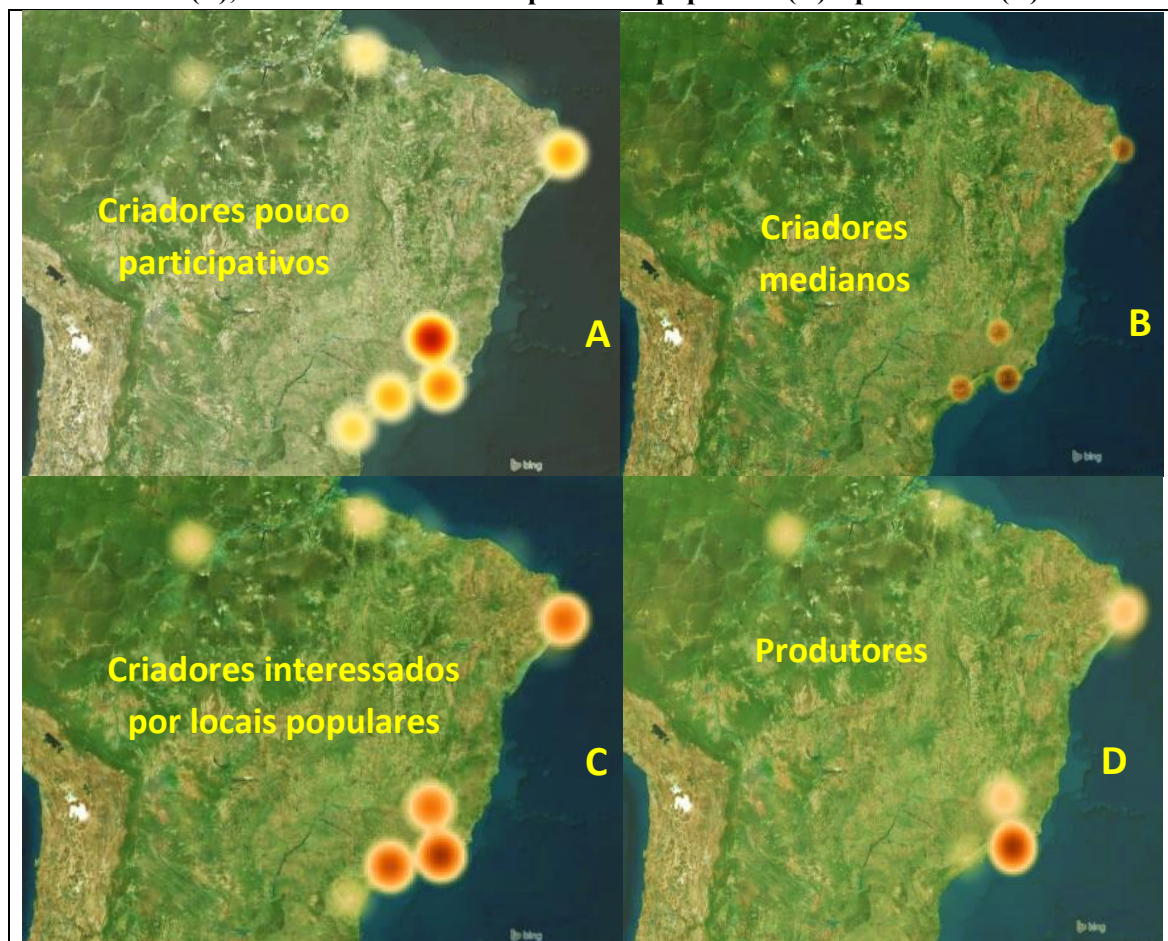
Figura 11. Distribuição dos gêneros por grupo de criador de conteúdo nas RGs



Fonte: Própria Autora.

Sobre a distribuição das participações dos grupos de colaboradores pelas cidades brasileiras, existe variações entre as regiões do Brasil que são vista na Figura 12. Mais especificamente, algumas cidades da região Sudeste do país concentram as participações de criadores pouco participativos (Figura 12A), criadores interessados por locais populares (Figura 12C) e produtores (Figura 12D). Sobre o compartilhamento de dicas nas regiões, a participação de criadores medianos é mais distribuída pelas regiões (Figura 12B).

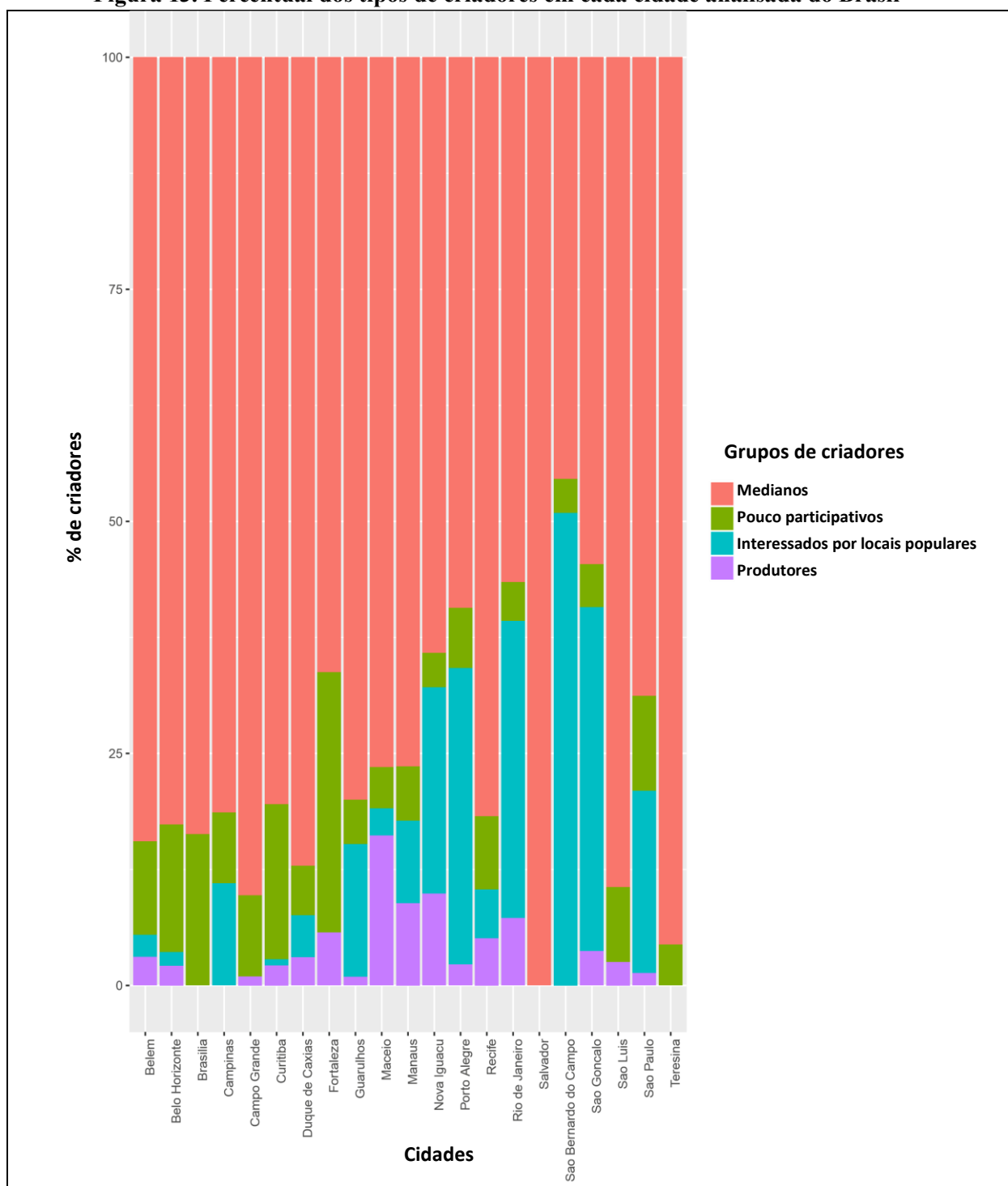
Figura 12. Intensidade da participação de criadores pouco participativos (A), criadores medianos (B), criadores interessados por locais populares (C) e produtores (D) no Brasil



Fonte: Própria Autora.

Outra observação diz respeito aos grupos de criadores por cidades. A Figura 13 e a Figura 14 (Apêndice C) indicam que a quantidade participativa dos perfis de criadores nas cidades é heterogênea. As proporções dos usuários por cidade que identifica que criadores medianos são maioria em quase todas as cidades brasileiras. Ainda nessa linha de observação, produtores e criadores pouco participativos são minoria nas cidades. As cidades de Manaus, Recife e Rio de Janeiro são compostas majoritariamente por produtores e ambas têm como característica serem capitais com uma população maior que 1,5 milhões. O perfil de criadores interessados por locais populares apresenta uma considerável incidência nas metrópoles do Rio de Janeiro e São Paulo, o que é compreensível por serem cidades com inúmeras opções de locais populosos. Por fim, a cidade de Salvador contém apenas um tipo de criador em toda a cidade (criadores medianos).

Figura 13. Percentual dos tipos de criadores em cada cidade analisada do Brasil



Fonte: Própria Autora.

A variedade na participação dos grupos, vista nas Figura 12 e Figura 13, indica que cada cidade apresenta particularidade sobre as proporções de tipos de criadores que coexistem. Talvez, a variação na participação dos colaboradores nas cidades seja motivada por razões externas à RG. Isso porque um mesmo criador de conteúdo pode interagir em várias cidades

com a RG, utilizando os mesmos recursos. Essa constatação dá origem à seguinte pergunta de pesquisa: “Quais características urbanas podem influenciar a participação dos criadores de conteúdo?”.

O Brasil é o décimo país mais desigual do mundo, segundo dados do Relatório de Desenvolvimento Humano (RDH), elaborado pelas Nações Unidas (ONU) (BREILLER, 2017). Como consequência, as disparidades entre as cidades brasileiras são existentes. Em outras palavras, há cidades brasileiras com ótimos indicadores urbanos (sobre segurança, infraestrutura, poder aquisitivo dos usuários, etc.) e outras cidades em que o comportamento é oposto.

Para responder o questionamento anterior, algumas características urbanas são selecionadas, como: a) Índice de Desenvolvimento Humano (IDH), b) população, c) número de assassinatos por armas de fogo, d) renda per capita, e) número de assassinatos femininos e f) população masculina. Tais fatores urbanos são escolhidos porque mensuram a qualidade de vida, a renda, a segurança e o porte populacional das cidades. Tais características são obtidas da base de dados do Instituto Brasileiro de Geografia e Estatística (IBGE).

Novamente, é aplicada o teste de correlação de Kendall e as saídas dessas execuções estão detalhadas nas Figura 15, Figura 16, Figura 17 e Figura 18 (Apêndice D). De acordo com os resultados obtidos, a maioria dos grupos de criadores tem seu padrão de comportamento ao contribuir afetado por fatores externos, relacionados às características urbanas.

Em cidades com IDH maior do que a média, criadores pouco participativos e medianos tendem compartilhar dicas em menos locais diferentes. Aparentemente, cidades com qualidade de vida melhor possuem locais que atraem a permanência das colaborações desses dois tipos criadores na RG.

Por outro lado, existem perfis de colaboradores que não têm o comportamento colaborativo influenciado por fatores urbanos. De acordo com os resultados da correlação de Kendall, criadores interessados por locais populares não são afetados pelas características urbanas selecionadas. Esse resultado cria duas possibilidades de justificativa. A primeira seria que as variáveis urbanas selecionadas na análise não são limitantes ou incentivadoras na busca por locais povoados de participação com a RG. Provavelmente, seria fatores relacionados à densidade populacional ou às áreas residenciais e comerciais da cidade. Outro motivo seria que criadores interessados em locais populares não são afetados por fatores urbanos, porque

esse tipo de colaborador já segue o fluxo da maioria dos participantes da RG que já é afetada por essas questões.

Por fim, produtores são o grupo de criadores de conteúdo que são mais afetados pelos fatores urbanos. Produtores tendem a colaborar em mais locais diferentes nas cidades mais populosas e mais violentas (taxa de homicídios por armas de fogo). Além do mais, eles tendem a preferir locais mais populares para compartilhar dicas em cidades com uma qualidade de vida melhor (IDH maior do que a média) e uma população com um alto poder aquisitivo (renda per capita maior do que a média).

5.7. Discussões dos resultados

Os resultados do mapeamento da diversidade dos colaboradores mostram que existem tipos de criadores de conteúdo nas cidades brasileiras, de acordo com os padrões de comportamento ao contribuir na RG. Um colaborador pode ter pouco interesse em criar conteúdo, como um criador pouco participativo; ou representar a grande maioria dos participantes, como os criadores medianos; ou preferir locais populosos para visitar, como criadores interessados por locais populares; ou, ainda, compartilhar informações ativamente, como os criadores produtores.

A análise de clusterização dá subsídios para constatações sobre a manutenção da RG. A primeira delas se refere aos criadores pouco participativos. Esse perfil de participante tem a conduta mais próxima aos *lurkers*: usuário típico de sistemas colaborativos que não colaboram com o sistema, apenas existe (GORIUNOVA, 2017). Baseado nisso, a propagação de criadores pouco colaborativos na comunidade pode comprometer o trabalho de analistas urbanos e administradores de sistemas, pois eles utilizam as interações na RG para tomar suas decisões. Além do mais, a elaboração de mecanismos de incentivo à colaboração é uma estratégia para distanciá-los da participação nula na comunidade.

Criadores medianos compõem o grupo mais equilibrado por causa de sua atitude em criar e consumir informação geolocalizada. Outra característica é que esse grupo confia nas avaliações dos lugares providas pela RG para compartilhar dicas. Assim, os administradores de sistemas devem elaborar mecanismos que encorajem a participação dos criadores medianos, como a exibição dos locais melhor avaliados numa área, quando este usuário estiver procurando locais para colaborar.

Criadores interessados por locais populares são detectores de zonas concentradas de

colaborações e essa característica é valiosa para analistas urbanos durante o mapeamento de áreas ativas da cidade.

Por fim, criadores produtores são os principais provedores de informação geolocalizada na comunidade da RG. Esse é o grupo que mais influencia a comunidade de colaboradores quando eles seguem seu próprio fluxo de colaborações. Além do mais, eles são importantes para as cidades devido à sua capacidade em descobrir novos lugares e por ser um modelo de colaborador a ser seguido, no quesito participação. Uma questão que também deve ser observada é que produtores são o tipo de colaborador mais sensível aos fatores urbanos. Isso quer dizer que dependendo das características da cidade, sua contribuição na RG pode ser prejudicada ou incentivada.

Todos os grupos de criadores percorrem uma distância curta para colaborar com a RG. Um fato importante é que o presente experimento detecta que as características urbanas utilizadas não são responsáveis por este comportamento dos criadores de conteúdo. No caso dos criadores interessados por locais populares, provavelmente é porque eles buscam por um tipo de lugar que é ponto de concentração de outros criadores e um local popular normalmente está próximo de outros. Criadores produtores percorrem as menores distâncias provavelmente para compartilhar a maior quantidade de colaborações dentre todos os grupos.

5.8. Conclusões do mapeamento

O presente experimento mostra que o principal desafio de investigação sobre a diversidade dos criadores de conteúdo diz respeito à seleção coerente das métricas que caracterizam os padrões de comportamento ao contribuir. A investigação do comportamento colaborativo abre novas perspectivas sobre os perfis de participantes, presentes em diferentes contextos urbanos. Entender a diversidade dos criadores de conteúdo pode melhorar mecanismos usados nas escolhas dos lugares na RG, tais como recomendação e filtragem de informações personalizadas.

Cada grupo tem sua importância para a comunidade. Primeiramente, criadores produtores representam o modelo ideal de colaborador na RG. Entretanto, eles são pouco receptivos às informações compartilhadas por terceiros. Além do mais, a presença de produtores é essencial para geração de conteúdos e para a elaboração de estratégias capazes de transformar outros criadores em produtores.

Os resultados indicam que o principal receptor de informações geolocalizada são os

criadores medianos e eles apresentam um equilíbrio entre a aquisição e provimento das informações geolocalizadas na RG. Criadores interessados por locais populares são os “detectores” dos locais muito frequentados, mas em áreas reduzidas.

O último grupo é o criador pouco participativo, composto de colaboradores com um baixo nível de interação com a RG. Baseado nisso, os sistemas geolocalizados devem elaborar maneiras de distanciar esses colaboradores do comportamento dos *lurkers* na comunidade.

Por fim, uma característica marcante entre os perfis de criadores se refere à curta distância percorrida por eles durante a interação com a RG. Nesse sentido, surgem novos questionamentos:

- Locais populares se concentram nas mesmas áreas das cidades brasileiras e, por isso, os perfis de criadores percorrem curtas distâncias para colaborar na RG?
- As cidades observadas apresentam opções limitadas de locais que obrigam os criadores de conteúdo a se comportarem como um mesmo tipo de perfil?

Isso será investigado no capítulo seguinte.

6 DIVERSIDADE BASEADA NO COMPORTAMENTO ESPACIAL

Esta fase do experimento mapeia a diversidade dos criadores de conteúdo, frente ao comportamento espacial nas cidades brasileiras. Em outras palavras, é observado os padrões espaciais nas interações dos participantes da RG em áreas pré-definidas.

Segundo os resultados do experimento anterior, a maioria dos criadores de informações geolocalizadas percorrem trajetórias reduzidas entre os locais visitados. Além do mais, não está claro as razões pelos quais eles desbravam um curto perímetro das cidades. A impressão é que, talvez, cidades brasileiras apresentem limitações na oferta de locais ou contenham áreas concentradas de interações com a RG. Baseado nisso, as seguintes perguntas de pesquisa são elaboradas:

- Como é a distribuição das colaborações realizadas pelos criadores de conteúdo em cidades brasileiras?
- Quais são padrões espaciais das colaborações dos criadores de conteúdo em cidades brasileiras?

6.1. Aspectos metodológicos

O presente experimento executa um estudo exploratório a respeito da diversidade espacial dos colaboradores de conteúdo geolocalizado. As interações observadas e compartilhadas sobre os locais de cidades brasileiras são no formato de *check-ins*. A escolha de *check-ins* foi motivada por ser uma perspectiva de colaboração que também valida o mapeamento espacial dos contribuidores.

Um criador u_1 compartilha um *check-in* no local l_1 , onde o *check-in* é definido como a tupla $\langle u_1, l_1 \rangle$. No local l_1 , outros n criadores de conteúdo também colaboram com a RG, $\langle u_2, l_1 \rangle$, $\langle u_3, l_1 \rangle$, ..., $\langle u_n, l_1 \rangle$. Assim, o presente trabalho foca no conjunto de tuplas relacionadas aos *check-ins* de RGs sobre diversos locais presentes nas cidades brasileiras.

A análise espacial dos criadores acontece por meio da observação de categorias de cidades brasileiras, baseada: i) na região da cidade, ii) no Índice de Desenvolvimento Humano (IDH) da cidade e iii) no gênero majoritário da cidade. A escolha dessas três características serve para detalhar aspectos sobre qualidade de vida, localidade espacial e característica

demográfica das cidades selecionadas. Por fim, a investigação dos criadores de conteúdo por grupo de cidades é importante para avaliar se, efetivamente, as propriedades urbanas definem comportamentos espaciais distintos entre os colaboradores de RGs.

A amostra dos dados utilizada contém *check-ins* feitos em 25 cidades brasileiras (Águas de São Pedro, Ananindeua, Balneário Camboriú, Belém, Belo Horizonte, Brasília, Caruaru, Campina Grande, Florianópolis, Fortaleza, Guareí, Iaras, João Pessoa, Manaus, Niterói, Olinda, Recife, Reginópolis, Rio de Janeiro, Salvador, Santos, São Pedro de Alcântara, São Caetano do Sul, São Paulo e Vitória) entre março e julho de 2014 na RG *Swarm*. As cidades escolhidas são representantes das categorias de cidades que apresentam as maiores quantidades de *check-ins* coletados.

A amostra analisada contém 1.901.102 *check-ins* em 25 cidades. Cada uma das cidades apresenta uma média de 82.683 *check-ins*, compartilhados por 56.585 criadores de conteúdo. A cidade com mais colaborações é São Paulo (371.818 *check-ins*) e a cidade com menos interações é Iaras (30 *check-ins*), ambas localizadas no estado de São Paulo.

Em média, cada colaborador compartilha 27,16 *check-ins*. Além disso, a cidade com a maior taxa de colaboração é João Pessoa com 76,05 *check-ins* por criador de conteúdo e a cidade com a menor taxa de colaboração foi Iaras com 3 *check-ins* por colaborador. A seguir, serão descritos os resultados de acordo com as categorias de cidades, citadas anteriormente.

A análise espacial dos criadores de conteúdo envolve a observação das participações e a análise de clusterização nos setores censitários das cidades selecionadas.

A observação da participação dos colaboradores nas cidades utiliza a técnica de autocorrelação espacial, com o índice de Moran. Esse método representa a forma básica de índices globais de autocorrelação espacial com similaridade entre as unidades de áreas (CLIFF e ORD, 1970). A autocorrelação espacial refere-se ao agrupamento de fenômenos semelhantes e díspares, que sinaliza a provável existência de subgrupos. Assim, esse método é responsável por representar o comportamento predominante dos setores censitários de cada cidade, em termos de participação na RG.

A análise de clusterização das colaborações utiliza duas técnicas principais: a análise de pontos de concentração (*hot spot*) e a análise de *clusters* e *outliers*. A análise de pontos de concentração (*hot spot*) utiliza a estatística de Getis-Ord G_i^* para cada colaboração do criador de conteúdo (ANSELIN, 1995). Além disso, a observação dos *z-scores* e dos *p-valores* resultantes indica onde os recursos com valores altos ou baixos se agrupam espacialmente.

Esse tipo de análise sobre a densidade das interações com a RG utiliza vetores para identificar as localidades com alta e baixa concentração.

O outro método para fins de clusterização é a análise de *clusters* e *outliers*, com o índice de Anselin Local Moran. Esse método verifica os *outliers* espaciais das colaborações, além de detectar os *clusters* espaciais de acordo com as concentrações das colaborações dos criadores nas cidades (ANSELIN, 1995).

Todo o mapeamento espacial observa as interações dos criadores em áreas menores das cidades, chamadas de setores censitários. Esse tipo de delimitação territorial é definido pelo IBGE para fins de controle cadastral da população. Com relação à instrumentação, *scripts* em Python e a ferramenta ArcGIS são utilizados para executar os métodos descritos anteriormente.

6.2. Resultados obtidos

Esta seção contém os detalhes obtidos na observação da atividade espacial dos colaboradores de conteúdo geolocalizado nos três cenários apresentados anteriormente.

6.2.1. *Diversidade do comportamento espacial de criadores de conteúdo, por região das cidades brasileiras*

O Brasil apresenta cinco regiões, de acordo com as similaridades físicas, culturais e econômicas das cidades. A ideia dessa fase do experimento é observar como é a participação espacial dos criadores de conteúdo em cada uma das regiões do Brasil.

O panorama das contribuições dos criadores de conteúdo nas cidades de cada região pode ser visto no Apêndice E e no Quadro 4. Eles descrevem os resultados da autocorrelação espacial, baseada no índice de Moran. O Quadro 4 resume o comportamento espacial dos colaboradores nas cidades selecionadas, detalhado no Apêndice E. Os resultados do índice de Moran indicam que cidades de uma mesma região podem conter colaboradores com comportamentos espaciais variantes. Por exemplo, os criadores de conteúdo podem estar concentrados ou não em áreas das cidades, independentemente da região a qual pertencem, como aparenta nas distribuições do Apêndice F. Portanto, a autocorrelação espacial detecta a ausência de similaridade entre os colaboradores, em relação ao comportamento espacial global de cada cidade.

Embora os colaboradores não apresentem padrões espaciais por região, ainda existe a possibilidade de existir similaridades locais na participação deles em áreas urbanas. Assim, o

passo seguinte é entender como as áreas visitadas pelos criadores de conteúdo estão agrupadas localmente nas cidades de cada região. O primeiro método utilizado é o método de análise de *clusters* e *outliers*, utilizando o índice de Anselin Local Moran. O resultado de cada uma das cidades pode ser visto no Apêndice G.

De acordo com os resultados do Quadro 8, cada área urbana pode assumir um dos seguintes tipos, baseados na intensidade da participação de criadores na RG: a) áreas de alta colaboração, rodeadas por outros setores intensamente colaborados (HH), b) áreas de baixa colaboração, próximas de outros setores com uma baixa criação de conteúdo (LL), c) área de alta taxa de participação, rodeados por outros setores com baixa colaboração (HL) e d) setores de baixa colaboração, rodeados por outros setores com alta criação de conteúdo (LH).

Quadro 4. Comportamento espacial das colaborações dos criadores de conteúdo, de acordo com a autocorrelação espacial

REGIÃO	CIDADE	COMPORTAMENTO ESPACIAL	P-VALOR
NORDESTE	Fortaleza	Aleatório	-
	Campina Grande	Aleatório	-
	João Pessoa	<i>Agrupado</i>	< 0,005
	Caruaru	<i>Agrupado</i>	< 0,005
	Recife	<i>Agrupado</i>	< 0,005
	Salvador	<i>Agrupado</i>	< 0,005
	Olinda	<i>Agrupado</i>	< 0,005
SUL	Balneário Camboriú	<i>Agrupado</i>	< 0,005
	Florianópolis	Aleatório	-
	São Pedro de Alcântara	Aleatório	-
CENTRO-OESTE	Brasília	<i>Agrupado</i>	< 0,005
SUDESTE	Vitória	Aleatório	-
	Águas de São Pedro	Aleatório	-
	São Caetano do Sul	<i>Agrupado</i>	< 0,005
	Belo Horizonte	<i>Agrupado</i>	< 0,005
	Guareí	Aleatório	-
	Santos	Aleatório	-
	São Paulo	<i>Agrupado</i>	< 0,005
	Rio de Janeiro	<i>Agrupado</i>	< 0,005
	Iaras	<i>Agrupado</i>	< 0,005
	Reginópolis	Aleatório	-
	Niterói	<i>Agrupado</i>	< 0,005
NORTE	Manaus	Aleatório	-
	Ananindeua	Aleatório	-
	Belém	<i>Agrupado</i>	< 0,005

Fonte: Própria Autora.

Os mapas oferecem subsídios para analisar a espacialização da participação dos criadores em diferentes níveis e regiões. Nas cidades do Nordeste, a maioria dos criadores preferem colaborar nas mesmas áreas das cidades. A prova disso é a detecção de vários setores com alta criação de conteúdo (setores HH e HL do Quadro 8). Uma minoria dos criadores de conteúdo nas cidades de Caruaru, Recife e Olinda também frequentam setores pouco populares (setores LH do Quadro 8).

Além do mais, há regiões que apresentam cidades com poucas áreas concentradas de participações dos colaboradores, como é o caso das cidades das regiões Sul. Por fim, não há similaridade espacial nas interações dos criadores nas regiões Sudeste, Norte e Centro-Oeste.

A última etapa de observação do comportamento espacial das interações nas áreas urbanas diz respeito à análise de pontos de concentração de colaboradores. Esses pontos podem

ter níveis de intensidade altos e baixos e segue a estatística de Getis-Ord G_i^* .

As descrições desse método podem ser vistas em mapas do Quadro 9 (Apêndice H). De acordo com os resultados expostos nos mapas, as cidades observadas da região Nordeste apresentam áreas de concentração de alta e baixa intensidade. Além do mais, não há um padrão de comportamento entre os pontos de concentração das áreas urbanas nas regiões Sul, Centro-Oeste e Norte. Entretanto, grandes centros da região Sudeste apresentam muitas áreas urbanas com alta e baixa concentração na criação de conteúdo geolocalizado. Outro fato interessante é que a cidade de Manaus apresenta apenas áreas urbanas com baixa taxa de concentração, ao contrário das cidades de sua região que também apresentam pontos de alta concentração.

6.2.2. Diversidade do comportamento espacial de criadores de conteúdo, baseado no IDH

De acordo com a ONU⁸, IDH é uma medida com dimensões sobre o desenvolvimento humano: quanto maior esse índice, maior é a chance de a população ligada à localidade ter uma vida longa e saudável, ser experiente e ter um padrão decente de vida. O IDH é uma média geométrica dos índices normalizados para cada uma das três dimensões.

O presente trabalho observa o comportamento espacial das cinco cidades com os maiores IDHs do Brasil: São Caetano do Sul, Águas de São Pedro, Florianópolis, Balneário Camboriú e Vitória. O objetivo dessa etapa da análise espacial é entender se a qualidade de vida define padrões na diversidade espacial das colaborações de criadores de conteúdo.

Inicialmente, a autocorrelação espacial mostra que não há uma padronização na participação global dos criadores de conteúdo em cidades com os melhores IDH do Brasil. Os detalhes desses resultados estão expostos no Apêndice E. As cidades brasileiras com os melhores indicadores de IDH não são unânimes quanto ao comportamento geral dos colaboradores nas cidades. Existem cidades, como São Caetano do Sul e Balneário Camboriú, em que os criadores de conteúdo se agrupam nas mesmas áreas urbanas e o restante delas em que há um padrão aleatório.

Num segundo momento, a análise de similaridades entre as áreas das cidades utiliza a técnica de análise de *clusters* e *outliers*. De acordo com os resultados do Apêndice G, as cidades selecionadas, exceto Águas de São Pedro, contêm áreas da cidade em que os criadores de

⁸ Disponível em: <http://hdr.undp.org/en/content/human-development-indexhdi>

conteúdo colaboram massivamente e são rodeados por setores pouco visitados por eles.

As cidades com os melhores IDH do Brasil apresentam pontos de alta concentração de colaborações de criadores de conteúdo, exceto Vitória. Novamente, detalhes sobre os pontos de concentração das interações estão no Apêndice H. Vale ressaltar que a cidade de Florianópolis também apresenta pontos de concentração com baixa intensidade de participação dos colaboradores na RG.

6.2.3. Diversidade do comportamento espacial de criadores de conteúdo, baseado no gênero majoritário da cidade

Nesta etapa da análise é levado em consideração se o gênero predominante das cidades determina padrões espaciais específicos para os criadores de conteúdo. As principais cidades brasileiras majoritariamente femininas são: Santos (54,25% de mulheres na população), Recife (53,87% de mulheres na população), São Caetano do Sul (53,85% de mulheres na população), Olinda (53,76% de mulheres na população) e Niterói (53,69% de mulheres na população). Além disso, as principais cidades masculinas são: Iaras (65,81% de homens na população), Reginópolis (63,93% de homens na população), São Pedro Alcântara (63,57% de homens na população) e Guareí (61,25% de homens na população) (GLOBO, 2010).

Em relação à análise espacial na observação geral, a maioria das cidades femininas (Recife, São Caetano do Sul, Olinda e Niterói) contém criadores de conteúdo que se agrupam nas áreas urbanas para compartilhar *check-ins*. Nas cidades masculinas, apenas a cidade de Iaras apresenta colaboradores que interagem com a RG em localidades próximas, criando um comportamento agrupado na participação da cidade. Novamente, todas as descrições dos resultados estão no Apêndice E.

No que se refere à análise local da participação dos criadores de conteúdo, o método de detecção de *clusters* e *outliers* detecta que os colaboradores são responsáveis por compartilhar informações geolocalizadas em áreas altamente concentradas nas cidades femininas. E nas cidades masculinas, os criadores de conteúdo não apresentam um padrão espacial de participação nas áreas urbanas. Aparentemente, não existe setores urbanos com uma ampla preferência dos criadores de conteúdo geolocalizado nas cidades masculinas.

Em relação aos pontos de concentração, é detectado que a maioria das cidades femininas e masculinas apresentam pontos de concentração com alta e baixa intensidade de participação dos criadores de conteúdo, com exceção das cidades de Guareí, Reginópolis e Iaras.

6.3. Discussão do mapeamento

Neste experimento, alguns padrões espaciais foram detectados entre os colaboradores da RG. Os conteúdos compartilhados podem estar concentrados ou dispersos em áreas urbanas e distribuídos de acordo com cada cidade observada. Embora a diversidade espacial já tenha sido detectada em estudos prévios com RGs em outros países (YANG *et al.*, 2016; YANG *et al.*, 2017), este mapeamento representa a primeira comprovação da diversidade espacial entre os criadores de conteúdo geolocalizado nas cidades brasileiras.

Outro ponto marcante das análises se refere às categorias de cidades que não se mostram determinantes no comportamento espacial dos colaboradores. A utilização de grupos de cidades busca por similaridades entre os colaboradores de RG, mas tais semelhanças não foram tão expressivas quanto esperado.

Quase todas as cidades observadas possuem criadores de conteúdo que se deslocam para áreas populares a fim de compartilhar informações geolocalizadas. Isso demonstra uma preferência comum entre determinados colaboradores. Nesse sentido, algumas justificativas surgem sobre a concentração de *check-ins* em áreas específicas das cidades. Esse comportamento pode ser motivado por:

- a) Características urbanas das áreas visitadas;
- b) Mecanismos dentro da RG, que podem incentivar os colaboradores a compartilharem *check-ins* em localidades próximas.

Com relação à primeira suspeita, é investigada a participação espacial nos setores censitários e se ela é influenciada pelos fatores urbanos dessas áreas. Para responder isso, é aplicada a análise de correlação de Kendall entre a intensidade das colaborações em cada setor e suas respectivas características urbanas. As características urbanas são descrições sobre o tipo de setor censitário, os tipos de domicílios contidos no setor, o poder aquisitivo dos moradores de cada setor, a infraestrutura de cada setor e detalhes sobre os moradores dos setores. Detalhes sobre cada uma das variáveis selecionadas no Quadro 10 (Apêndice I).

Os resultados das correlações estão expostos no Quadro 11 (Apêndice J) e mostram que criadores de conteúdo variam a intensidade de suas colaborações na RG, de acordo com as características urbanas de algumas cidades. Isso significa que cidades como Guareí, Águas de São Pedro, Reginópolis e João Pessoa possuem criadores de conteúdo que compartilham mais intensamente em áreas residenciais. Outra descoberta mostra que as cidades de Brasília,

Caruaru, Campina Grande, Fortaleza, Guareí, João Pessoa, Recife e São Pedro de Alcântara possuem criadores de conteúdo que se concentram mais em áreas com um alto poder aquisitivo dos moradores.

Em contrapartida, a maioria das cidades (Ananindeua, Belém, Balneário Camboriú, Belo Horizonte, Florianópolis, Manaus, Niterói, Olinda, Rio de Janeiro, Salvador, Santos, São Caetano do Sul, São Paulo, Serra Azul e Vitória) possui colaboradores que tem sua interação na RG afetada por nenhum fator urbano.

A outra possibilidade sobre a participação concentrada em áreas específicas das cidades é devido aos mecanismos dentro da RG. No caso da RG Foursquare quando um usuário faz *check-in* em uma localidade, o sistema provê para ele algumas informações sobre o local, como qual dos seus amigos já estiveram no local, quantos *check-ins* já foram feitos no lugar e a nota atribuída ao lugar. Locais bem visitados, bem avaliados e com muitos amigos frequentadores apresentam uma probabilidade maior do criador de conteúdo ter interesse em colaborar também.

A resposta dos questionamentos pode dar subsídios para novos mecanismos dentro da RG que sejam sensíveis à localidade do usuário. Por exemplo, se o mesmo estiver em áreas concentradas das cidades, o próprio sistema poderia sugerir locais em áreas próximas com características similares para uma possível interação.

7 CONCLUSÃO

Os três mapeamentos executados mostram que a diversidade dos colaboradores de RGs pode ser descrita sob diversas facetas e este trabalho explicitou três delas, que são referentes à experiência sobre lugares, ao perfil colaborativo e ao comportamento espacial. A escolha dessas abordagens foi baseada nas variáveis amplamente utilizadas nos trabalhos correlatos, além de representar a diversidade do criador de conteúdo com conhecimentos, ações colaborativas e hábitos espaciais variados. Dessa forma, cada um dos momentos do mapeamento tem relevância para o entendimento da diversidade dos criadores de conteúdo geolocalizado.

7.1. Sumário

O presente trabalho tem como objetivo solucionar a ausência de estudos sobre a diversidade de criadores de conteúdo geolocalizado em cidades brasileiras. A primeira dessas soluções diz respeito à análise da variedade dos colaboradores, baseada na experiência sobre locais durante um evento turístico na cidade. Esse experimento tem como principal contribuição a descoberta que criadores de conteúdo com mais conhecimento sobre locais da cidade são essencialmente mais relevantes para a comunidade. A variedade de colaboradores, de acordo a experiência, já havia sido discutida em RGs de outros países (HAN *et al.*, 2014), mas no Brasil essa descoberta é uma novidade importante.

A segunda observação se refere à investigação da diversidade dos colaboradores frente aos padrões de comportamento ao contribuir nas cidades brasileiras. Nesta análise, fica evidente que embora o público de cada cidade tenha uma dinâmica particular com a RG, existem tipos de criadores de conteúdo que são comuns na maioria delas. Além disso, os padrões de comportamento ao contribuir dos colaboradores são influenciados diferentemente pelas características socioeconômicas da cidade.

Um direcionamento comum aos dois primeiros mapeamentos diz respeito à investigação do impacto de características urbanas sobre a diversidade dos colaboradores. Trabalhos na literatura (EAGLE, MACY e CLAXTON, 2010; JIN *et al.*, 2016) mostram como esse questionamento é pertinente, pois enriquece a personalização dos usuários nesses sistemas. Nos mapeamentos, as indagações sobre os fatores urbanos surgiram ao longo da execução do experimento e após a elaboração das perguntas de pesquisa.

A influência das características urbanas na interação dos participantes da RG levou a criação da última investigação da diversidade sob a ótica do comportamento espacial. Neste experimento é observado o comportamento espacial dos colaboradores durante as interações. Os resultados das análises mostram que colaboradores de RGs apresentam alguns padrões espaciais ao colaborar em determinadas cidades brasileiras.

7.2. Contribuições para a pesquisa em RGs

A principal contribuição do presente trabalho é a análise em larga-escala da variedade dos criadores de conteúdo geolocalizado em cidades brasileiras. Além disso, a proposta de um processo metodológico bem documentado e reproduzível. Estudos destacam as RGs como uma fonte relevante na investigação sobre as dinâmicas espaço-temporais de áreas urbanas (BALDUINI *et al.*, 2014; CRANSHAW *et al.*, 2012; MCKENZIE *et al.*, 2015; NOULAS *et al.*, 2012; WANG *et al.*, 2014). Além disso, as RGs são úteis durante a elaboração de soluções modernas da análise urbana (ZHENG *et al.*, 2014), através da descrição das propriedades espaciais das atividades que ocorrem dentro de uma área. Por fim, esses mapeamentos incluíram o Brasil na lista de países com alta receptividade de RGs, que tiveram seus usuários analisados isoladamente.

Os resultados apresentados abrem novas possibilidades de investigação dos criadores de conteúdo. Uma delas se refere à análise da dinamicidade da diversidade dos criadores de conteúdo na RG. Isso significa a elaboração de mecanismos que “rastreiem” o colaborador a fim de avaliar se ele muda de comportamento ao longo do tempo e da localidade. Todas as variáveis utilizadas nos mapeamentos da presente tese podem ser utilizadas para esse suporte.

7.3. Contribuições para o *design* de RGs

Os resultados obtidos sobre a diversidade dos contribuidores servem de subsídio para diversos mecanismos internos às RGs. O primeiro deles se refere às técnicas de recomendação de informações. Recomendações são técnicas nas RGs responsáveis por sugerir locais para os usuários da comunidade. Normalmente, a recomendação é baseada em características do participante (PARK, HONG e CHO, 2017), no histórico colaborativo do criador de conteúdo (TAKEUCHI e SUGIMOTO, 2006) ou até mesmo de acordo com o grupo de amigos virtuais (HSU *et al.*, 2006).

Os resultados sobre a diversidade possibilitam uma dinamicidade nas técnicas recomendações porque podem se basear nos comportamentos dos criadores de conteúdo relacionados à experiência, aos padrões de comportamento ao contribuir e aos hábitos espaciais. Um exemplo concreto seria a melhoria no processo de recomendação usando a filtragem colaborativa. Nesse processo, as recomendações são feitas a partir da correlação entre os usuários da RG. Se a filtragem das informações for realizada pela correlação do tipo de usuário, a especificidade poderia ser aumentada.

Um segundo exemplo seria as técnicas de recomendação que utilizam mineração de dados para descobrir o conhecimento dos usuários. A análise da diversidade auxiliaria no registro do conhecimento específico para cada tipo de colaborador. Portanto, seria necessário saber qual tipo de conhecimento o participante apresenta na comunidade.

Outro mecanismo que se beneficia com o mapeamento da diversidade é a personalização nas RGs. Como mencionado em seções anteriores, a quase totalidade das RGs tratam seus usuários homogeneamente. O grande desafio da personalização é prover mecanismo que não sejam tão complexos na análise individual. Num primeiro momento, é muito dispendioso prover recursos individualmente nas RGs. Os resultados sobre a diversidade são importantes para simplificar a personalização porque contém informações úteis para previsão de recursos de acordo com o tipo de colaborador, sem necessariamente criar uma RG diferente para cada participante.

A compreensão sobre diversidade dos criadores de conteúdo também serve de subsídio para a proposição de melhorias na tomada de decisão de acordo com o tipo de usuário. Por exemplo, a elaboração de estratégias dentro das RGs que incentivem a participação dos usuários pouco participativos.

Segundo Yang *et al.* (2016), a diversidade representa um desafio e uma oportunidade para experimentos e sistemas de análises urbanas. Mecanismos, como a mineração de dados, também podem se apropriar das análises sobre a diversidade para auxiliar os analistas nas suas decisões.

7.4. Beneficiados pela pesquisa

A temática deste trabalho não beneficia apenas os usuários das RGs. O mapeamento da diversidade entre os colaboradores é útil para diversos agentes. Primeiramente, ela é útil para proprietários de locais públicos que podem compreender quais são os tipos de visitantes que

frequentam sua propriedade. Em segundo lugar, a diversidade é importante para desenvolvedores de RGs definirem que recursos são realmente válidos para o público-alvo. E, por fim, é útil para pesquisadores avaliarem a diversidade em análises multidisciplinares.

As RGs desempenham um importante papel na obtenção detalhada dos retratos das cidades devido aos seus dados em larga escala e à ausência de ferramentas mais apropriadas. Nesse sentido, a importância da utilização de RGs para analistas urbanos tem sido cada vez maior.

Vale ressaltar também que a metodologia executada não limita o estudo da diversidade apenas para criadores de conteúdo. Isso significa que o processo metodológico usado para mapear a diversidade beneficia pesquisadores que investigam usuários de RGs em contextos urbanos pouco explorados, como África e América Latina, por exemplo.

Outros favorecidos com os resultados sobre a diversidade dos colaboradores são áreas fora da computação. O primeiro exemplo seria a área de urbanismo, que pode observar o fluxo de movimentação de diferentes tipos de colaboradores pela cidade. Com esse controle das trajetórias, é possível detectar problemas de mobilidade, áreas de repulsão da população ou até mesmo elaborar políticas de urbanização das áreas pouco frequentadas nas cidades.

Outra área que se beneficia com a observação da diversidade em RGs é a área turística. Após compreender que turistas e residentes não vivenciam os lugares públicos da mesma maneira durante um evento, políticas de apoio a esses dois grupos poderiam ser elaboradas separadamente.

7.5. Publicações

Outro resultado relevante do trabalho diz respeito às publicações da presente pesquisa na comunidade científica. A primeira delas é sobre a análise da diversidade dos criadores em relação à experiência num trabalho intitulado de “*The Relevance of Annotations Shared by Tourists and Residents on a Geo-Social Network during a Large-Scale Touristic Event: the case of São João*” no *11th International Conference on the Design of Cooperative Systems* em 2014 (MORAIS e ANDRADE, 2014). Em seguida, uma análise sobre o comportamento espacial dos colaboradores de Campina Grande é publicada no *International Journal on Web Service Computing* num trabalho intitulado como “*Spatial Analysis About Users Collaboration On Geo-Social Networks In A Brazilian City*” (MORAIS, 2015). Ainda no ano de 2015, também foi publicado um resumo estendido intitulado “*The Diversity of Users on*

Geo-Social Networks” no *4th GESIS Summer School in Survey Methodology* (MORAIS e ANDRADE, 2015).

No ano de 2016 foi publicado a análise da diversidade frente ao comportamento espacial em várias cidades brasileiras nas Conferências Ibero-Americanas WWW/Internet e Computação Aplicada em um trabalho intitulado “Diversidade Espacial dos Usuários em Redes Geossociais: Estudo de Caso em Cidades Brasileiras” (MORAIS, 2016). Por fim, o detalhamento da análise da diversidade em relação aos padrões de comportamento ao contribuir dos participantes está publicado no trabalho “*Profiles of Mapping of Content Creators in a Geo-Social Network: The Case of 21 Brazilian Cities*” em *First Monday - peer reviewed Journal on the Internet* (MORAIS e ANDRADE, 2017).

7.6. Trabalhos futuros

Há continuidade na linha de pesquisa desempenhada por esta tese. A primeira delas se refere ao desenvolvimento de uma arquitetura de RGs, orientada à diversidade dos criadores de conteúdo. Dependendo do tipo de escopo da RG, a arquitetura pode conter todas ou algumas das abordagens relacionadas à diversidade.

A segunda frente de futuros trabalhos é a comparação da diversidade dos criadores de conteúdo do Brasil com colaboradores de outros países. O intuito seria mapear se criadores de conteúdo brasileiros possuem algum comportamento peculiar na RG. As descobertas obtidas nos mapeamentos não podem ser comparadas com resultados de outros estudos já realizados até o momento. Primeiramente porque eles não foram executados seguindo o mesmo processo metodológico. Segundo, porque a natureza da diversidade não se assemelha entre os estudos já realizados.

Por fim, outra linha de investigação da diversidade é a análise temporal da diversidade dos criadores. A dinamicidade dos padrões de comportamento ao contribuir, da experiência e comportamento espacial dos colaboradores das RGs é existente. Mas, ainda não sabemos como é essa variação numa escala temporal.

7.7. Considerações finais

De acordo com Liu *et al.* (2010), anunciantes interessam-se por sistemas colaborativos com variedade de usuários porque concentram participantes de diferentes tipos que podem ser

clientes em potencial. O Brasil é um país sinônimo de diversidade fora dos sistemas colaborativos. Assim, analisar se essa variedade também acontece nas RGs do país é fundamental para a compreensão do Brasil como consumidor de sistemas colaborativos.

Os mapeamentos mostram quão variada é a forma que criadores de conteúdo percebem, experimentam e interagem com a comunidade de RG. Outro fato importante é que as abordagens da diversidade podem ser complementares entre si. Isso significa que podem existir nas RGs turistas que também são produtores e só colaboram em áreas bastante visitadas, como também turistas pouco participativos que preferem áreas pouco frequentadas, por exemplo.

Os resultados sobre criadores de conteúdo podem parecer específicos ao contexto observado nos experimentos (cidades brasileiras em vários escopos: por temporadas do ano, com um bom IDH, as mais populosas, etc.). Na verdade, essas descobertas dão auxílio para RGs lidarem com criadores de conteúdos que não foram analisados nesses mapeamentos, mas estão colaborando em cidades brasileiras com características similares às apresentadas.

Com relação à replicação dos experimentos, alguns desafios devem ser expostos. Primeiramente, os mecanismos de obtenção de dados podem apresentar limites no número de requisições obtidas. A observação de uma quantidade massiva de colaborações só foi possível porque a amostra detém interações desde 2012. Outro desafio diz respeito à utilização de métricas que necessariamente precisam obedecer aos cálculos pré-definidos pela pesquisa.

As discussões realizadas indicam que a diversidade dos colaboradores não é uma característica estática. Por exemplo, um criador de conteúdo geolocalizado nem sempre será um residente na comunidade geolocalizada. Portanto, a diversidade não deve ser encarada como uma variável estática inserida nos mecanismos de personalização. Na verdade, a diversidade deve ser vista como um critério dinâmico de observação dos criadores de conteúdo.

BIBLIOGRAFIA

ANSELIN, L. Local indicators of spatial association—LISA. **Geographical analysis**, v. 27, n. 2, p. 93-115, 1995.

ANSELIN, L.; WILLIAMS, S. Digital neighborhoods. **Journal of Urbanism: International Research on Placemaking and Urban Sustainability**, v. 9, n. 4, p. 305-328, 2016.

ARNABOLDI, V.; CONTI, M.; DELMASTRO, F. Implementation of CAMEO: A context-aware middleware for Opportunistic Mobile Social Networks. **World of Wireless, Mobile and Multimedia Networks (WoWMoM)**, 2011 IEEE International Symposium on a. IEEE, 2011, p. 1-3.

BALDUINI, M. *et al.* Recommending venues using continuous predictive social media analytics. **IEEE Internet Computing**, v. 18, n. 5, p. 28-35, 2014.

BALLESTEROS, J. *et al.* Towards safe cities: A mobile and social networking approach. **IEEE Transactions on Parallel and Distributed Systems**, v. 25, n. 9, p. 2451-2462, 2014.

BAUER, C.; STRAUSS, C. Location-based advertising on mobile devices. **Management Review Quarterly**, v. 66, n. 3, p. 159-194, 2016.

BENJAMIN, D. J. *et al.* Redefine statistical significance. **Nature Human Behaviour**, p. 1, 2017.

BHOLOWALIA, P.; KUMAR, A. EBK-means: A clustering technique based on elbow method and k-means in WSN. **International Journal of Computer Applications**, v. 105, n. 9, 2014.

BHUIYAN, T. Trust for Intelligent Recommendation - **Springer Briefs in Electrical and Computer Engineering**. Editora Springer, 2013, p. 31-32. ISBN 1461468957.

BOYD, D.; CRAWFORD, K. Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. **Information, Communication & Society**, v. 15, n. 5, p. 662-679, 2012.

BREILLER P. EL PAÍS. Brasil despenca 19 posições em ranking de desigualdade social da ONU: País aparece entre os 10 mais desiguais do mundo. Além da diferença entre ricos e pobres, levantamento ressalta desvalorização e baixa representatividade da mulher na sociedade brasileira. Disponível em: <https://brasil.elpais.com/brasil/2017/03/21/politica/1490112229_963711.html>. Acesso em: 22 dez. 2017.

BROWN, C. *et al.* Group colocation behavior in technological social networks. **PloS one**, v. 9, n. 8, p. e105816, 2014.

CARRIGY, Tara *et al.* Design and evaluation of player experience of a location-based

mobile game. **Proceedings of the 6th Nordic conference on human-computer interaction: Extending boundaries**. ACM, 2010, p. 92-101.

CHANG, J.; SUN, E. Location 3: How users share and respond to location-based data on social networking sites. **Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media**. 2011, p. 74-80.

CHEN, C. *et al.* TripPlanner: Personalized trip planning leveraging heterogeneous crowdsourced digital footprints. **IEEE Transactions on Intelligent Transportation Systems**, v. 16, n. 3, p. 1259-1273, 2015.

CHEN, M. *et al.* Extreme learning machine based point-of-interest recommendation in location-based social networks. **Proceedings of ELM-2015**, v. 2. Springer, Cham, 2016, p. 249-261.

CHO, E.; MYERS, S.A.; LESKOVEC, J. Friendship and mobility: user movement in location-based social networks. **Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining**. ACM, 2011, p. 1082-1090.

CHON, K. S. Tourism in Southeast Asia: A new direction. **Editora Routledge**, 2013, p. 128-134.

CHORLEY, M. J.; WHITAKER, R. M.; ALLEN, S. M. Personality and location-based social networks. **Computers in Human Behavior**, v. 46, p. 45-56, 2015.

CLIFF, A. D.; ORD, K. Spatial autocorrelation: a review of existing and new measures with applications. **Economic Geography**, v. 46, n. sup1, p. 269-292, 1970.

CRANSHAW, J. *et al.* The livelihoods project: Utilizing social media to understand the dynamics of a city. 2012.

CROUX, C.; DEHON, C. Influence functions of the Spearman and Kendall correlation measures. **Statistical methods & applications**, v. 19, n. 4, p. 497-515, 2010.

DEL BIMBO, A.; FERRACANI, A.; PEZZATINI, D. Flarty: recommending art routes using check-ins latent topics. **Proceedings of the 21st ACM international conference on Multimedia**. ACM, 2013, p. 457-458.

DMR - BUSINESS STATISTICS. 20 important foursquare stats and facts. Disponível em: <https://expandedramblings.com/index.php/by-the-numbers-interesting-foursquare-user-stats/>. Acesso em: 20 dez. 2017.

DUTZ, T. *et al.* User interfaces of mobile exergames. **International Conference on Human-Computer Interaction**. Springer, Cham, 2014, p. 244-255.

EAGLE, N.; MACY, M.; CLAXTON, R. Network diversity and economic development. **Science**, v. 328, n. 5981, p. 1029-1031, 2010.

EMARKETER. **Argentina Enjoys Highest Mobile Social User Penetration in Latin America – Brazil home to largest mobile phone social network audience in the**

region. Disponível em: <https://www.emarketer.com/Article/Argentina-Enjoys-Highest-Mobile-Social-User-Penetration-Latin-America/1011998>. Acesso em: 20 dez. 2017.

EXAME. **Foursquare cresceu 800% no brasil em 12 meses.** Disponível em: https://exame.abril.com.br/tecnologia/foursquare-cresceu-800-no-brasil-nos-ultimos-12-meses/?utm_source=blog&utm_campaign=rc_blogpost. Acesso em: 20 dez. 2017.

FARRAR, D. E.; GLAUBER, R. R. Multicollinearity in regression analysis: the problem revisited. **The Review of Economic and Statistics**, p. 92-107, 1967.

GAO, H.; LIU, H. Data analysis on location-based social networks. **Mobile social networking**. Springer New York, 2014, p. 165-194.

GEORGIEV, P.; NOULAS, A.; MASCOLO, C. The call of the crowd: Event participation in location-based social services. AAAI. 2014.

GIACOMELE, S. PmWeb. **Relatório/Resumo: A Internet no Brasil em 2015.** Disponível em: <https://blog.pmweb.com.br/a-internet-no-brasil-em-2015/>. Acesso em: 20 dez. 2017.

GLOBO. **Nove dos dez municípios com mais homens estão em SP.** Disponível em: <http://g1.globo.com/brasil/noticia/2010/11/nove-dos-dezmunicipios-com-mais-homens-estao-em-sp-diz-ibge.html>. Acesso em: 20 dez. 2017.

GORIUNOVA, O. Global Digital Culture| The Lurker and the Politics of Knowledge in Data Culture. **International Journal of Communication**, v. 11, p. 17, 2017.

GUO, B. *et al.* TaskMe: toward a dynamic and quality-enhanced incentive mechanism for mobile crowd sensing. **International Journal of Human-Computer Studies**, v. 102, p. 14-26, 2017.

HAN, K. *et al.* Enhancing community awareness of and participation in local heritage with a mobile application. **Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing**. ACM, 2014, p. 1144-1155.

HARDY, J.; LINDTNER, S. Constructing a Desiring User: Discourse, Rurality, and Design in Location-Based Social Networks. **Proceedings of Computer Supported Cooperative Work (CSCW)**, 2017, p. 13-25.

HARLOW, L. L.; MULAİK, S. A.; STEIGER, J. H. What if there were no significance tests?: classic edition. **Ed. Routledge**, 2016.

HAWELKA, B. *et al.* Geo-located Twitter as proxy for global mobility patterns. **Cartography and Geographic Information Science**, v. 41, n. 3, p. 260-271, 2014.

HIGGINS, J. P. *et al.* Measuring inconsistency in meta-analyses. **BMJ: British Medical Journal**, v. 327, n. 7414, p. 557, 2003.

HJORTH, L.; WILKEN, R.; GU, K. Ambient intimacy: A case study of the iPhone, presence, and location-based social media in Shanghai, China. **Routledge**, 2012, p. 43-62.

HO, Y. H.; WU, Y. C.; CHEN, M. C. PLASH: a platform for location aware services with human computation. **IEEE Communications Magazine**, v. 48, n. 12, p. 44-51, 2010.

HONIG, C.D.F.; MACDOWALL, L. Spatio-temporal mapping of street art using Instagram. **First Monday**, v. 22, n. 3, 2017. DOI: <http://dx.doi.org/10.5210/fm.v22i3.7072>.

HRISTOVA, D. *et al.* Measuring urban social diversity using interconnected geo-social networks. **Proceedings of the 25th International Conference on World Wide Web**. International World Wide Web Conferences Steering Committee, 2016, p. 21-30.

HSU, W. H. *et al.* Collaborative and Structural Recommendation of Friends using Weblog-based Social Network Analysis. **AAAI Spring Symposium: Computational Approaches to Analyzing Weblogs**. 2006, p. 55-60.

HUANG, A.; GALLEGOS, L.; LERMAN, K. Travel analytics: Understanding how destination choice and business clusters are connected based on social media data. **Transportation Research Part C: Emerging Technologies**, v. 77, p. 245-256, 2017.

HUMPHREYS, L.; LIAO, T. Mobile geotagging: Reexamining our interactions with urban space. **Journal of Computer-Mediated Communication**, v. 16, n. 3, p. 407-423, 2011.

JAIN, A. K. **Data clustering: 50 years beyond K-means**. Pattern recognition letters, v. 31, n. 8, p. 651-666, 2010.

JAMISON-POWELL, S. *et al.* Understanding in-situ social media use at music festivals. **Proceedings of the Companion Publication of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing**. ACM, 2014. p. 177-180.

JERÔNIMO, C. L. M.; CAMPELO, C. E.C.; DE SOUZA BAPTISTA, C. Using Open Data to Analyze Urban Mobility from Social Networks. **Journal of Information and Data Management**, v. 8, n. 1, p. 83, 2017.

JIN, L. *et al.* Characterizing users' check-in activities using their scores in a location-based social network. **Multimedia Systems**, v. 22, n. 1, p. 87-98, 2016.

KANG, J.; LERMAN, K. Leveraging user diversity to harvest knowledge on the social web. **Third International Conference on Social Computing (SocialCom)**, IEEE, 2011, p. 215-222.

KAPLAN, E. L.; MEIER, P. Nonparametric estimation from incomplete observations. **Journal of the American Statistical Association**, v. 53, n. 282, p. 457-481, 1958.

KONG, L.; LIU, Z.; HUANG, Y. Spot: Locating social media users based on social network context. **Proceedings of the VLDB Endowment**, v. 7, n. 13, p. 1681-1684, 2014.

KONOMI, S. Beyond mobile collaboration: toward metropolitan-scale geocentric crowdsourcing. **Proceedings of the 2nd International Workshop on Ubiquitous Crowdsourcing**. ACM, 2011, p. 17-20.

LACIC, E. *et al.* Utilizing Online Social Network and Location-Based Data to Recommend Items in an Online Marketplace. **arXiv preprint arXiv:1405.1837**, 2014.

LEE, M.; CHUNG, C. A user similarity calculation based on the location for social network services. **International Conference on Database Systems for Advanced Applications**. Springer, Berlin, Heidelberg, 2011, p. 38-52.

LI, J. *et al.* Where to place the next outlet? harnessing cross-space urban data for multi-scale chain store recommendation. **Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct**. ACM, 2016, p. 149-152.

LI, R. *et al.* Co-occurrence prediction in a large location-based social network. **Frontiers of Computer Science**, v. 7, n. 2, p. 185-194, 2013.

LI, Y. *et al.* Dissecting foursquare venue popularity via random region sampling. **Proceedings of the 2012 ACM conference on CoNEXT student workshop**. ACM, 2012, p. 21-22.

LI, Y.; LI, Q.; SHAN, J. Discover patterns and mobility of Twitter users—A study of four US college cities. **ISPRS International Journal of Geo-Information**, v. 6, n. 2, p. 42, 2017.

LIANG, Y. *et al.* How big is the crowd?: event and location based population modeling in social media. **Proceedings of the 24th ACM Conference on Hypertext and Social Media**. ACM, 2013, p. 99-108.

LINDQVIST, J. *et al.* I'm the mayor of my house: examining why people use foursquare—a social-driven location sharing application. **Proceedings of the SIGCHI Conference on Human Factors in Computing Systems**. ACM, 2011, p. 2409-2418.

LINS, T.; PEREIRA, A. C.M.; BENEVENUTO, F. Workload characterization of a location-based social network. **Social Network Analysis and Mining**, v. 4, n. 1, p. 209, 2014.

LONG, X.; JIN, L.; JOSHI, J. Exploring trajectory-driven local geographic topics in foursquare. **Proceedings of the 2012 ACM Conference on Ubiquitous Computing**. ACM, 2012, p. 927-934.

MCKENZIE, G. *et al.* How where is when? On the regional variability and resolution of geosocial temporal signatures for points of interest. **Computers, Environment and Urban Systems**, v. 54, p. 336-346, 2015.

MCPHERSON, M.; SMITH-LOVIN, L.; COOK, J. M. Birds of a feather: Homophily in social networks. **Annual Review of Sociology**, v. 27, n. 1, p. 415-444, 2001.

MOHAMED, S.; ABDELMOTY, A. I. Spatio-semantic user profiles in location-based social networks. **International Journal of Data Science and Analytics**, p. 1-16, 2017.

MORAIS, A. Diversidade espacial dos usuários em redes geossociais: estudo de caso em cidades brasileiras. Conferência Ibero-Americanas WWW/Internet e Computação Aplicada, 2016, p.141-148. ISBN: 978-989-8533-59-3.

MORAIS, A. M.; ANDRADE, N. Profiles of mapping of content creators in a geo-social network: The case of 21 Brazilian cities. **First Monday**, v. 22, n. 12, 2017.

MORAIS, A. Spatial analysis about users collaboration on geo-social networks in a Brazilian city. **International Journal on Web Service Computing (IJWSC)**, 2015, v.6, n.4, p.1-13.

MORAIS, A.; ANDRADE, N. The Diversity of Users on Geo-Social Networks. **Computational Social Science Winter Symposium**. 2015. Resumo estendido.

MORAIS, A.; ANDRADE, N. The Relevance of Annotations Shared by Tourists and Residents on a Geo-Social Network During a Large-Scale Touristic Event: The Case of São João. **Proceedings of the 11th International Conference on the Design of Cooperative Systems**, 27-30 May 2014, Nice (France). Springer, Cham, 2014, p. 393-408.

MUANGNA, U.; NETRAMAI, C. Social network for Thailand medical tourism. **Biomedical Engineering International Conference (BMEiCON)**, 2013 6th. IEEE, 2013, p. 1-5.

NASH, A.; PURGATHOFER, P.; KAYALI, F. Using Online Games in Transport: **Grr--Grr--Bike Case Study**. 2013.

NORMAN, G.; KESHA, N. Using smartphones for cycle planning. **IPENZ Transportation Group Conference**. 2015, p. 22-24.

NOULAS, A. *et al.* A tale of many cities: universal patterns in human urban mobility. **PloS one**, v. 7, n. 5, p. e37027, 2012.

NOULAS, A. *et al.* Exploiting Semantic Annotations for Clustering Geographic Areas and Users in Location-based Social Networks. **The Social Mobile Web**, v. 11, n. 2, 2011.

O'CONNOR, M.; HERLOCKER, J. Clustering items for collaborative filtering. **Proceedings of the ACM SIGIR Workshop on Recommender Systems**. UC Berkeley, 1999.

O'CONNOR, P. Managing a hotel's image on TripAdvisor. **Journal of Hospitality Marketing & Management**, v. 19, n. 7, p. 754-772, 2010.

PARK, M.; HONG, J.; CHO, S. Location-based recommendation system using bayesian user's preference model in mobile devices. **International Conference on Ubiquitous Intelligence and Computing**. Springer, Berlin, Heidelberg, 2007, p. 1130-1139.

PARK, S. *et al.* Identical user tracking with behavior pattern analysis in online community. **Proceedings of the Symposium on Applied Computing**. ACM, 2017. p. 1086-1089.

PAVLAKIS, P.; ALEPIS, E.; VIRVOU, M. Intelligent mobile multimedia application for the support of the elderly. **Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP)**, 2012 Eighth International Conference on. IEEE, 2012. p. 297-

PERKINS, L. What Is Community Management?. **The Community Manager's Playbook**. Apress, 2015, p. 3-20.

PREOȚIUC-PIETRO, D.; COHN, T.. Mining user behaviours: a study of check-in patterns in location based social networks. **Proceedings of the 5th Annual ACM Web Science Conference**. ACM, 2013, p. 306-315.

RAHIMI, S. M.; WANG, X. Location recommendation based on periodicity of human activities and location categories. **Pacific-Asia Conference on Knowledge Discovery and Data Mining**. Springer, Berlin, Heidelberg, 2013, p. 377-389.

RÖSLER, R.; LIEBIG, T. Using data from location based social networks for urban activity clustering. **Geographic information science at the heart of Europe**. Springer International Publishing, 2013. p. 55-72.

ROUSSEEUW, P. J. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. **Journal of computational and applied mathematics**, v. 20, p. 53-65, 1987.

SAKER, M.; EVANS, L. Locative mobile media and time: Foursquare and technological memory. **First Monday**, v. 21, n. 2, 2016.

SALESSES, P.; SCHECHTNER, K.; HIDALGO, C. A. The collaborative image of the city: mapping the inequality of urban perception. **PloS one**, v. 8, n. 7, p. e68400, 2013.

SARKAR, D.; SIEBER, R.; SENGUPTA, R. GIScience Considerations in Spatial Social Networks. **International Conference on Geographic Information Science**. Springer International Publishing, 2016, p. 85-98.

SARWAT, M. *et al.* Plutus: leveraging location-based social networks to recommend potential customers to venues. Mobile Data Management (MDM), 2013 **IEEE 14th International Conference on**. IEEE, 2013, p. 26-35.

SARWAT, M. *et al.* The anatomy of Sindbad: a location-aware social networking system. **Proceedings of the 5th ACM SIGSPATIAL International Workshop on Location-Based Social Networks**. ACM, 2012, p. 1-4.

SCCELLATO, S.; MASCOLO, C. Measuring user activity on an online location-based social network. **Computer Communications Workshops (INFOCOM WKSHPS)**, 2011 IEEE Conference on. IEEE, 2011, p. 918-923.

SCHAAR, A. K.; VALDEZ, A. C.; ZIEFLE, M. The impact of user diversity on the willingness to disclose personal information in social network services. **Human Factors in Computing and Informatics**. Springer, Berlin, Heidelberg, 2013, p. 174-193.

SCHEFFE, H. The relation of control charts to analysis of variance and chi-square tests. **Journal of the American Statistical Association**, v. 42, n. 239, p. 425-431, 1947.

- SCOTT, J. **Social Network Analysis**. SAGE, 2017, 4^a ed., p. 6-7.
- SEBERT, D. M. Outliers in statistical data. **Journal of Quality Technology**, v. 29, n. 2, p. 230, 1997.
- SHANG, J. *et al.* A mobile guide system framework for museums based on local location-aware approach. **Computer Science and Service System (CSSS)**, 2011 International Conference on. IEEE, 2011, p. 1935-1940.
- SHARMA, A.; BAWA, R. K. A roadmap for agility estimation and method selection for secure agile development using AHP and ANN. **Data Engineering and Intelligent Computing**. Springer, Singapore, 2018, p. 237-245.
- SILVA, T. H. *et al.* A picture of Instagram is worth more than a thousand words: Workload characterization and application. **Distributed Computing in Sensor Systems (DCOSS)**, IEEE International Conference, 2013b, p. 123-132.
- SILVA, T. H. *et al.* Large-scale study of city dynamics and urban social behavior using participatory sensing. **IEEE Wireless Communications**, v. 21, n. 1, p. 42-51, 2014.
- SILVA, T. H. *et al.* Social media as a source of sensing to study city dynamics and urban social behavior: Approaches, models, and opportunities. **Ubiquitous Social Media Analysis**. Springer, Berlin, Heidelberg, 2013a, p. 63-87.
- SILVA, T. H. *et al.* Uma fotografia do Instagram: Caracterização e aplicação. **Revista Brasileira de Redes de Computadores e Sistemas Distribuídos**, 2017.
- SKARBERG, J.; SLETTEN, K.; KOLBEINSVIK, M. UTMOST: Unregistered Transport–Mobility, **Safety and new Technologies**. 2014.
- SMITH, J.; STRAIGHT, R.; FRANKLIN, T. Student Occupational Expectations: A Geolocate Study. **E-Learn: World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education**. Association for the Advancement of Computing in Education (AACE), 2011, p. 2522-2526.
- SOLIMAN, A. *et al.* Social sensing of urban land use based on analysis of Twitter users' mobility patterns. **PLoS One**, v. 12, n. 7, p. e0181657, 2017.
- SPOSARO, F.; DANIELSON, J.; TYSON, G. iWander: An Android application for dementia patients. **Engineering in Medicine and Biology Society (EMBC)**, 2010 Annual International Conference of the IEEE. IEEE, 2010, p. 3875-3878.
- SUITS, D. B. Use of dummy variables in regression equations. **Journal of the American Statistical Association**, v. 52, n. 280, p. 548-551, 1957.
- TAKEUCHI, Y.; SUGIMOTO, M. CityVoyager: an outdoor recommendation system based on user location history. **Ubiquitous Intelligence and Computing**, p. 625-636, 2006.
- TUNG, V. W. S.; RITCHIE, J. B. Exploring the essence of memorable tourism experiences. **Annals of Tourism Research**, v. 38, n. 4, p. 1367-1386, 2011.

VECCHIONE, A. E.; MELLINGER, M. Beyond FourSquare: library treks with SCVNGR. 2011.

WANG, H.; ZHU, L.; CHIN, A. An Indoor Location-based Social Network for Managing Office Resource and Connecting People. **Proceedings of the Symposia and Workshops on Ubiquitous, Autonomic and Trusted Computing**, 2010, p. 481-483.

WANG, Z. *et al.* Cross-domain community detection in heterogeneous social networks. **Personal and Ubiquitous Computing**, v. 18, n. 2, p. 369-383, 2014.

WILCOXON, F. Individual comparisons by ranking methods. **Biometrics bulletin**, v. 1, n. 6, p. 80-83, 1945.

WU, D.; MAMOULIS, N.; SHI, J. Clustering in Geo-Social Networks. **Bulletin of the IEEE Computer Society Technical Committee on Data Engineering**, 2015.

WU, F.; LI, Z. Where did you go: Personalized annotation of mobility records. **Proceedings of the 25th ACM International on Conference on Information and Knowledge Management**. ACM, 2016, p. 589-598.

XU, B.; DING, Z.; CHEN, H. Mining multiple periods in event time sequence. **Asia-Pacific Services Computing Conference**. Springer, Cham, 2015, p. 278-288.

YANG, J. *et al.* Diversity in urban social media analytics. **International Conference on Web Engineering**. Springer International Publishing, 2016, p. 335-353.

YANG, Z. *et al.* Indigenization of urban mobility. **Physica A: Statistical Mechanics and its Applications**, v. 469, p. 232-243, 2017.

YANO, M.; KAJI, K.; KAWAGUCHI, N. TimeTable. Locky: nation wide WiFi location information system based on user contributed information. **Proceedings of the 3rd International Workshop on Location and the Web**. ACM, 2010, p. 7.

YE, M. *et al.* On the Semantic Annotation of Places in Location-Based Social Networks. **Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'11)**, 2011, p. 520-528.

ZHENG, Y. *et al.* Urban computing: concepts, methodologies, and applications. **ACM Transactions on Intelligent Systems and Technology (TIST)**, v. 5, n. 3, p. 38, 2014.

ZHENG, Y. Tutorial on location-based social networks. **Proceedings of the 21st International Conference on World Wide Web, WWW**, 2012, v. 12, n. 5.

ZHENG, Y.; XIE, X.; MA, W. Geolife: A collaborative social networking service among user, location and trajectory. **IEEE Data Eng. Bull.**, v. 33, n. 2, p. 32-39, 2010.

ZHU, Y. *et al.* Predicting user activity level in social networks. **Proceedings of the 22nd ACM international conference on Information & Knowledge Management**. ACM, 2013, p. 159-168.

APÊNDICES

Apêndice A – Descrição das RGs selecionadas na busca sistemática

Neste apêndice é realizado o comparativo entre as RGs, considerando diferentes características. A primeira delas é sobre o segmento da cidade que é afetado pela RG urbana. Vale ressaltar que algumas RGs podem ter utilidade para mais de um segmento urbano. A segunda característica se refere ao público-alvo de cada RG. A maioria das RGs tem propósitos específicos para a cidade, por isso, elas são voltadas para um grupo específico de participantes.

A terceira característica utilizada no comparativo das RGs diz respeito ao impacto que elas causam no segmento da cidade. Isso se resume a uma breve explicação de como a RG funciona no segmento que ela atua. Em seguida, é verificado se a RG possui uma API de coleta ou de consulta. Essa característica é importante para garantir a reprodutibilidade em experimentos. Por fim, é relatada as vantagens e desvantagens de cada uma das RGs.

Quadro 5. RGs selecionadas em trabalhos científicos com impacto nas cidades

Rede Geossocial	Segmento da cidade	Público-alvo	Impactos no contexto urbano	Possui API?	Vantagem	Desvantagem
1. <i>Amazing Thailand</i>	Turismo	Turistas que estão na Tailândia	Informa ao turista passeios e experiências na Tailândia	Não	Ajuda usuários que não conhecem bem a Tailândia	Não complexidade na rede social
2. <i>Askus</i>	Urbanismo	Pessoas com dúvidas sobre localidades	Ajuda pessoas que não a resolver dúvidas sobre um local	Sim	Interface de Q&A geolocalizada	Interface visual apenas textual
3. <i>Broadcastr</i>	Entretenimento Turismo	Não há definição do público-alvo	Incentiva o aprendizado sobre a cidade por meio do registro de áudio de outros usuários sobre experiências nos locais	Não	Mídia não convencional utilizada nos registros dos usuários	A comunicação com a RG é assíncrona
4. <i>CAMEO</i>	Turismo	Turistas	Enriquece a experiência turística por meio de informações compartilhadas entre usuários, que podem não se conhecer (mas tem gostos parecidos)	Sim	Gera uma função de utilidade baseada no contexto para mensurar a relevância da informação	Funciona apenas em plataformas Android e possui comunicação assíncrona
5. <i>EKI.LOCKY</i>	Transporte	Pessoas que buscam informações sobre <i>wi-fi</i> e sobre linhas de trem	Permite que usuários saibam se a quanto tempo uma linha de trem está próxima do usuário	Não	Interação simples que mostra todas as tabelas com os tempos de respostas	Vinculado ao sistema IOS

6. <i>Facebook Places</i>	Entretenimento	Não há definição do público-alvo	Estimula a exploração e marcação de locais pela cidade	Sim	Relato sobre locais para seus contatos do Facebook	Apenas para mapeamento
7. <i>Find & Connect</i>	Comércio	Grupo de colegas de trabalho dentro de uma empresa	Estende os relacionamentos criados no trabalho para fora da empresa	Não	Facilita a organização de encontros de colegas dentro e fora da empresa	Vinculado ao sistema Symbian
8. <i>Flarty</i>	Turismo	Turistas em Florença	Informações e recomendações sobre Florença, Itália	Não	Promove a cidade	Específico para um tipo de cidade
9. <i>Foursquare</i>	Entretenimento Turismo Comércio	Não há definição do público-alvo	Os traços de jogabilidade permite a exploração pela cidade	Sim	Jogabilidade na interação com a RG	Não tem distinção entre turistas e residentes
10. <i>GeoLife</i>	Turismo	Viajantes	Melhora a experiência do viajante em locais que ele não conhece	Não	Constrói um histórico de experiências de acordo com a localização	Não há variedade de interações
11. <i>Gowalla</i>	Entretenimento Turismo Comércio	Pessoas que querem publicar suas informações sobre locais	Exploração do espaço urbano através do compartilhamento de suas localizações em uma rede social	Sim	Integrado com o Facebook	Caiu em desuso recentemente
12. <i>Instagram</i>	Entretenimento	Não há definição do público-alvo	Estimula a exploração e marcação de locais pela cidade por meio de fotos geolocalizadas	Não	Relato sobre a posição geolocalizada no Instagram	Não há muitos impactos no contexto urbano
13. <i>Intelligent Mobile Multimedia Application for the Support of the Elderly</i>	Saúde	Cuidadores de idosos, geriatras e idosos	Evita que um idoso se perca, pois manda alertas de sua localização para os cuidadores. Além de notificar prontamente o médico quando o cuidador relata algum alerta de saúde.	Não	Estabelece uma relação de independência funcional ente cuidador/idoso.	Não houve menção à família do paciente
14. <i>IWander</i>	Saúde	Portadores de alguma demência e cuidadores	Permite que pacientes e cuidadores tenham uma relação mais interdependente entre si, através do monitoramento espacial remoto	Não	Interação simples (por comando de voz ou texto) no módulo paciente	Vinculado ao sistema Android
15. <i>Lost State College</i>	Cultura	Qualquer usuário que queira partilhar ou aprender experiências sobre o legado de um local	Ajuda usuários na construção do legado de sua comunidade através de informações oficiais e do compartilhamento de experiências	Não	Promove o aprendizado sobre a consciência histórica de um local	Aplicado apenas no Nordeste dos EUA

16. <i>Mobile Guide System Framework for Museums</i>	Turismo	Visitantes de museus	Permite que usuários aprendam de forma independente de guias humanos sobre o museu	Sim	Utilização de diversas plataformas de comunicação que permitem diversos cenários de interação para o usuário	Não possui comunicação entre os usuários
17. <i>Moovit</i>	Transporte	Pessoas que utilizam transporte público	Informações sobre horários dos ônibus	Não	Informações atualizadas	Geolocalização assíncrona
18. <i>PLASH</i>	Urbanismo	Pessoas interessadas em resolver problemáticas urbanas em conjunto	Ajuda a solucionar questões relacionadas ao tráfego e ao panorama da cidade, por exemplo	Sim	É possível criar outros tipos de aplicações para o contexto urbano	Não tem um foco específico no contexto urbano
19. <i>PLUTUS</i>	Comércio	Donos de empreendimentos	O aplicativo atrai prováveis clientes para a empresa	Não	Ajuda usuários a visitarem novos locais comerciais da cidade	Finalidade empresarial
20. <i>Runking</i>	Saúde	Pessoas que praticam corrida	A RG compartilha o trajeto percorrido automaticamente para a rede de contatos	Não	Monitora os caminhos percorridos pelos usuários e ranqueia com seus contatos	Princípios simples de jogabilidade
21. <i>SCVNGR</i>	Entretenimento Turismo	Não há definição do público-alvo	Apresenta jogabilidade, que incentiva a exploração de locais da cidade	Sim	Valoriza a interação entre locais e usuários, por meio dos desafios	A geolocalização é opcional nesse sistema
22. <i>Sindbad</i>	Entretenimento Turismo	Usuários do Foursquare	Para qualquer usuário, há a recuperação do conteúdo compartilhado pelos amigos, de acordo com a localização do usuário	Sim	Melhoria das técnicas de recomendação dos RGs atuais (Foursquare, Facebook Places)	GUI um pouco ruim, comparada ao design dos RGs atuais
23. <i>Socialight</i>	Urbanismo	Usuários de uma cidade	Permite que o usuário expresse sua opinião sobre locais de uma cidade para a sua rede de amigos	Não	Troca de experiências entre usuários em uma cidade	Aplicado só em Nova Iorque
24. <i>SocialNTT</i>	Saúde	Turistas que precisam de informações médicas na Tailândia	Informa turistas sobre tipos de tratamentos disponíveis, hospitais mais próximos	Não	Também pode ser uma RG convencional	Se limita a Tailândia
25. <i>SocialTelescope</i>	Turismo	Turistas	Ranqueia os locais mais frequentados por meio de feedback de turistas	Não	Ranqueamento implícito	Não tem contribuição no contexto urbano em relação aos RGs convencionais
26. <i>TraMSNET</i>	Turismo	Turistas sem época específica	Prover informações para turistas de usuários que tenham as mesmas	Não	RG voltada para turistas	Não diferencia as experiências do turista.

			preferências e não necessariamente são amigos			
27. <i>TripAdvisor</i>	Turismo Comércio	Turistas	Possibilita que turistas obtenham e compartilhem informações sobre assuntos turísticos de uma cidade	Não	Troca de experiências entre turistas	Geolocalização é opcional
28. <i>Twitter</i>	Entretenimento	Não há definição do público-alvo	Estimula a exploração e marcação de locais pela cidade	Sim	Relato sobre a posição geolocalizada no Twitter	Apenas para mapeamento
29. <i>Viking Ghost Hunt</i>	Entretenimento	Jogadores de uma área	Jogo de caça-fantasma, que incentiva a exploração pela cidade	Não	Iniciativa recente sobre os jogos baseados em geolocalização	Explora características da localização, mas não agrega sobre o conhecimento da cidade
30. <i>Waze</i>	Transporte	Qualquer usuário que queira informações de trânsito ou solicitar carona	Popularização de redes efêmeras	Não	Informações em tempo real	Anonimato em algumas operações
31. <i>Whrrl</i>	Entretenimento Turismo	Não há definição do público-alvo	-	Não	-	Saiu de execução
32. <i>YELP</i>	Turismo	Qualquer usuário que deseja saber sobre um local	Avaliação sobre locais da cidade	Sim	Informações síncronas	Informações assíncronas
33. <i>Zaarly</i>	Comércio	Promove um bazar virtual para entre usuários de uma comunidade	Incentivo ao comércio entre usuários de uma mesma região	Não	Compradores não precisam conhecer previamente os vendedores e vice-versa	Não intermedia desavenças e nem devolve o dinheiro das transações comerciais

Fonte: Própria Autora.

Apêndice B – Linha do tempo das atividades descritas

Quadro 6. Linha do tempo das atividades por bimestre do ano

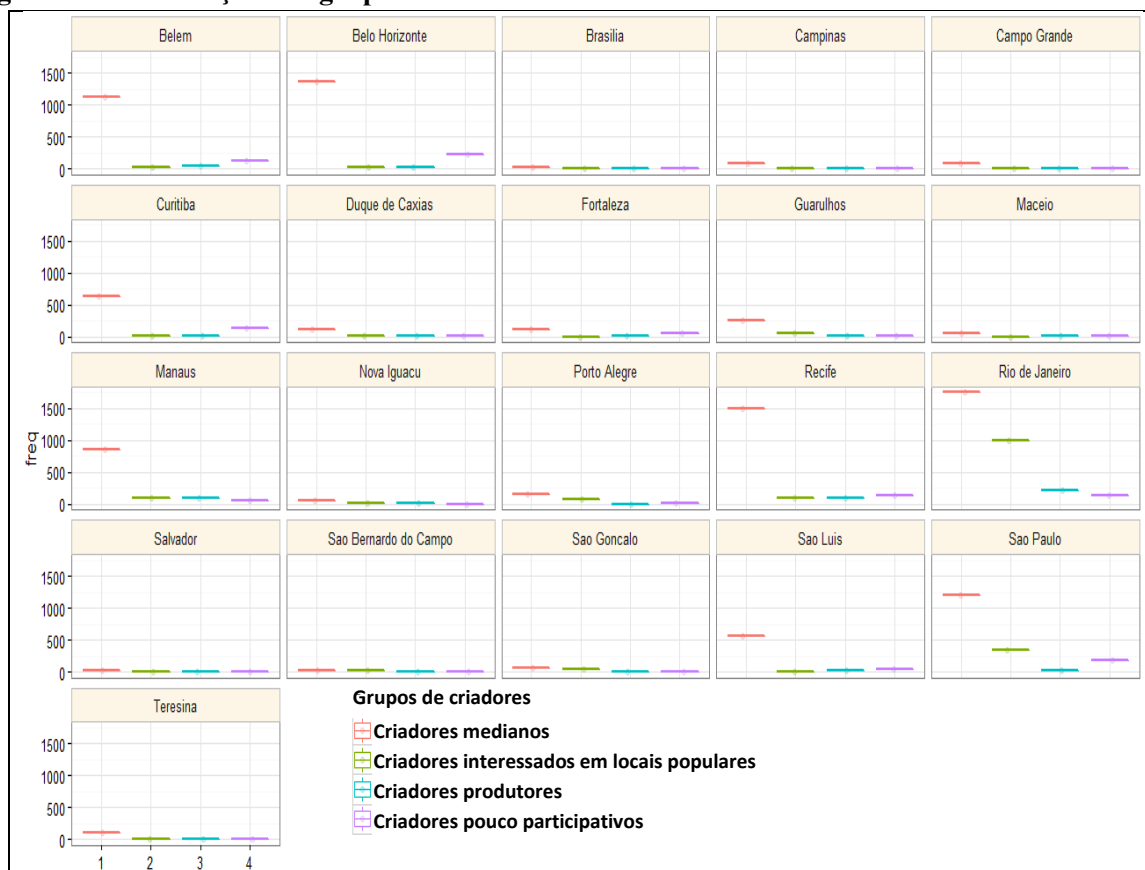
	2012				2013				2014				2015				2016				2017			
	I	II	III	IV	I	II	III	IV	I	II	III	IV	I	II	III	IV	I	II	III	IV	I	II	III	IV
<i>Investigação sobre RGs nas cidades</i>																								
<i>Investigação sobre a metodologia sobre a diversidade nas RGs</i>																								
<i>Investigação sobre as variáveis que caracterizam a diversidade dos criadores de conteúdo</i>																								
<i>Mapeamento da diversidade, frente à experiência</i>																								
<i>Mapeamento da diversidade, frente aos padrões de comportamento ao contribuir</i>																								
<i>Mapeamento da diversidade, frente ao comportamento colaborativo</i>																								

Fonte: Própria Autora.

Apêndice C – Participação dos perfis de criadores por cidade brasileira selecionada

Neste apêndice, é detalhado como os grupos de criadores de conteúdo se distribuem pelas cidades brasileiras selecionadas. De acordo com a Figura 14, o gráfico mostra como a quantidade de perfis de usuários são quantitativamente mais intensas em grandes cidades como as capitais Belém, Rio de Janeiro, São Paulo e Recife, por exemplo.

Figura 14. Distribuição dos grupos de criadores de conteúdo nas cidades observadas do Brasil



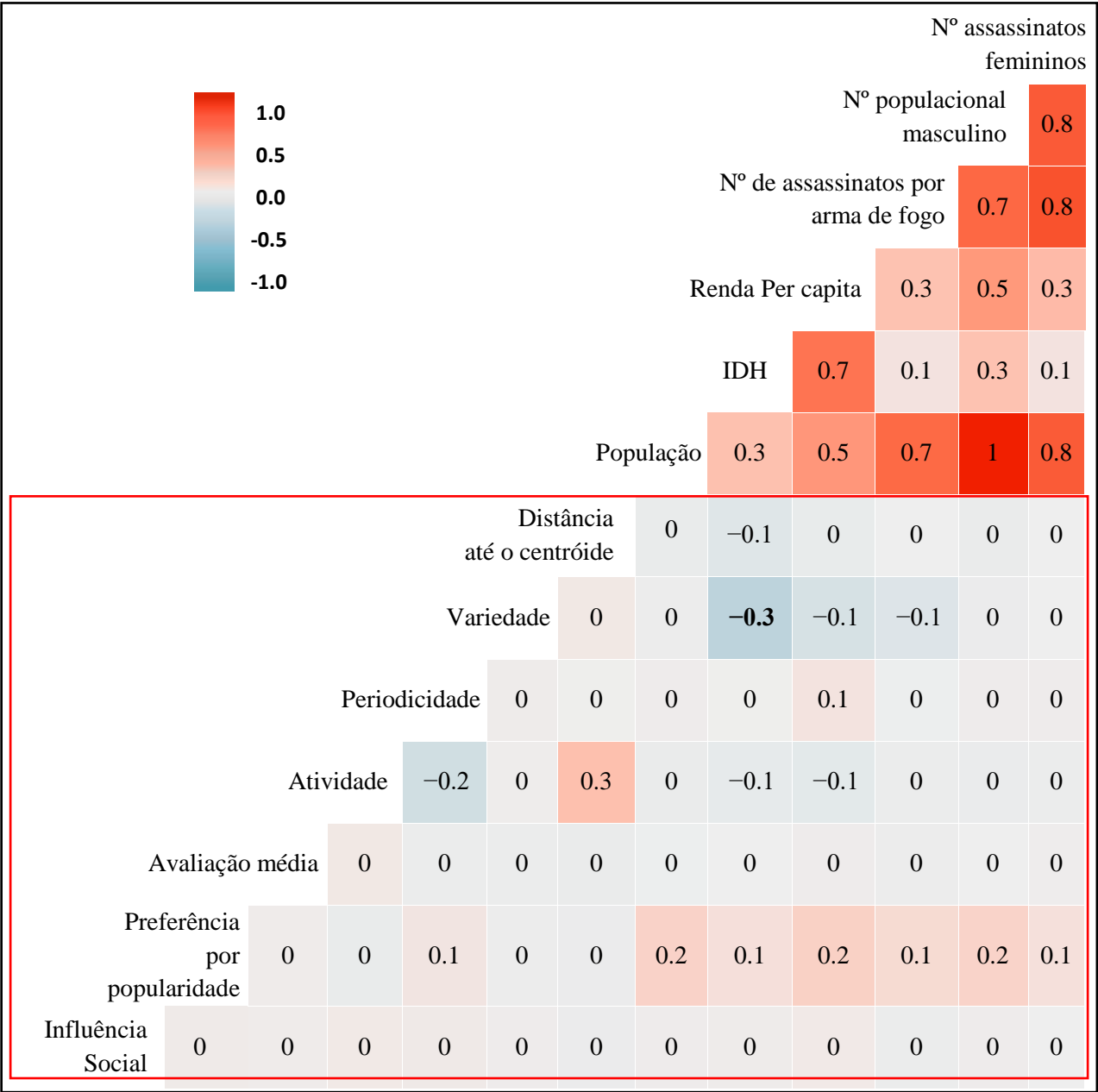
Fonte: Própria Autora.

Por outro lado, cidades como Salvador, São Bernardo do Campo e São Gonçalo, por exemplo, apresentam baixas ocorrências dos quatro grupos. Mais detalhes sobre as implicações desses resultados no Capítulo 7.

Apêndice D – Correlação entre características da participação colaborativa de criadores e as características urbanas das cidades brasileiras selecionadas

São apresentadas as análises de correlação de Kendall das características urbanas sobre os padrões de comportamento ao contribuir por grupo. A primeira delas está na Figura 15 e mostra que as principais influências são entre as características urbanas, mas não é o foco da pesquisa.

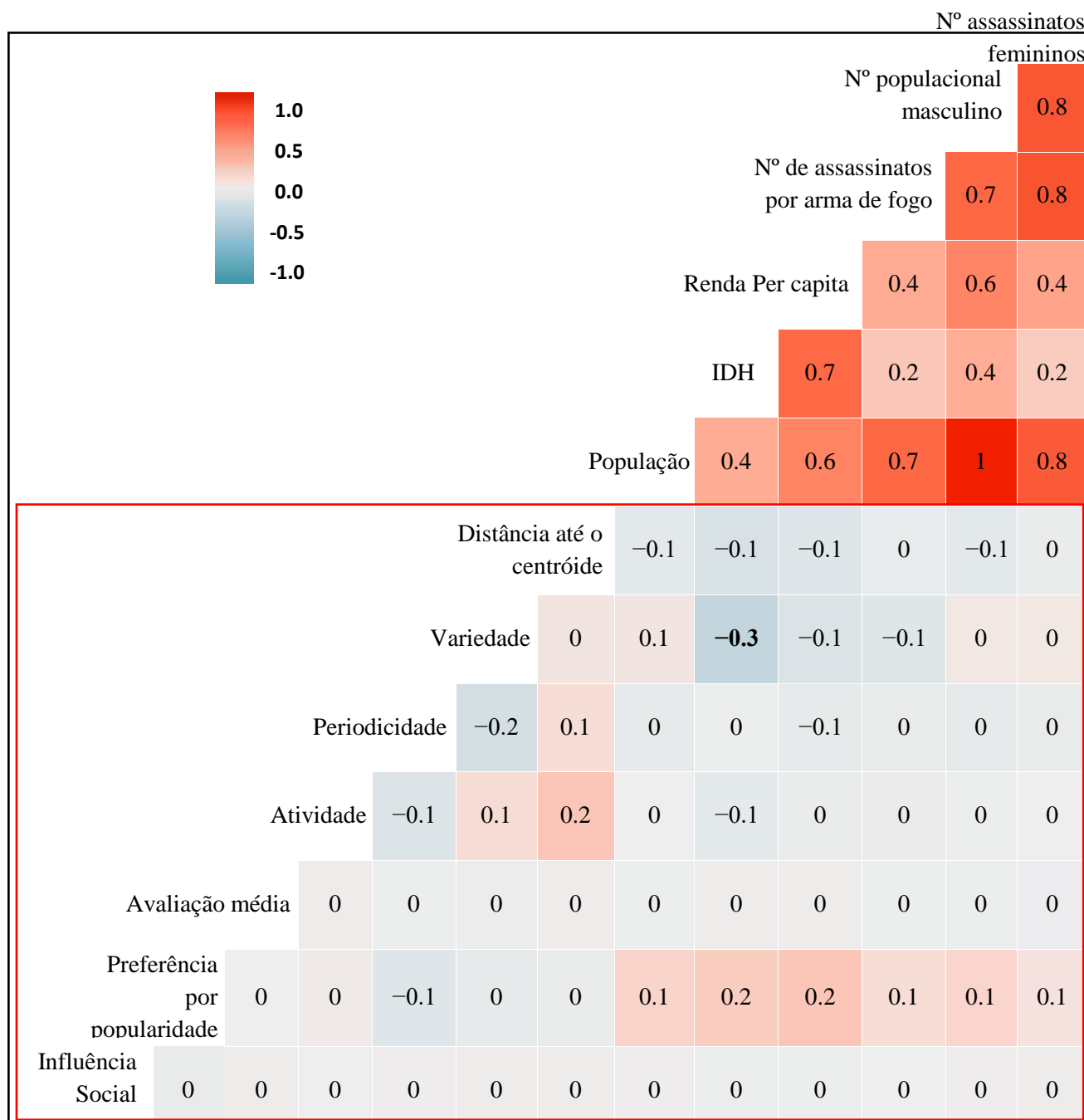
Figura 15. Correlação de Kendall da participação colaborativa dos criadores pouco participativos e as características urbanas das cidades



Fonte: Própria Autora.

Os resultados mostram que criadores pouco participativos tem a variedade afetada moderada e negativamente pelo IDH da cidade. Isso significa que eles variam mais de locais para colaborar em cidades com IDH mais baixo. A Figura 16 mostra a correlação de Kendall para os criadores medianos. Os resultados são similares aos criadores pouco participativos.

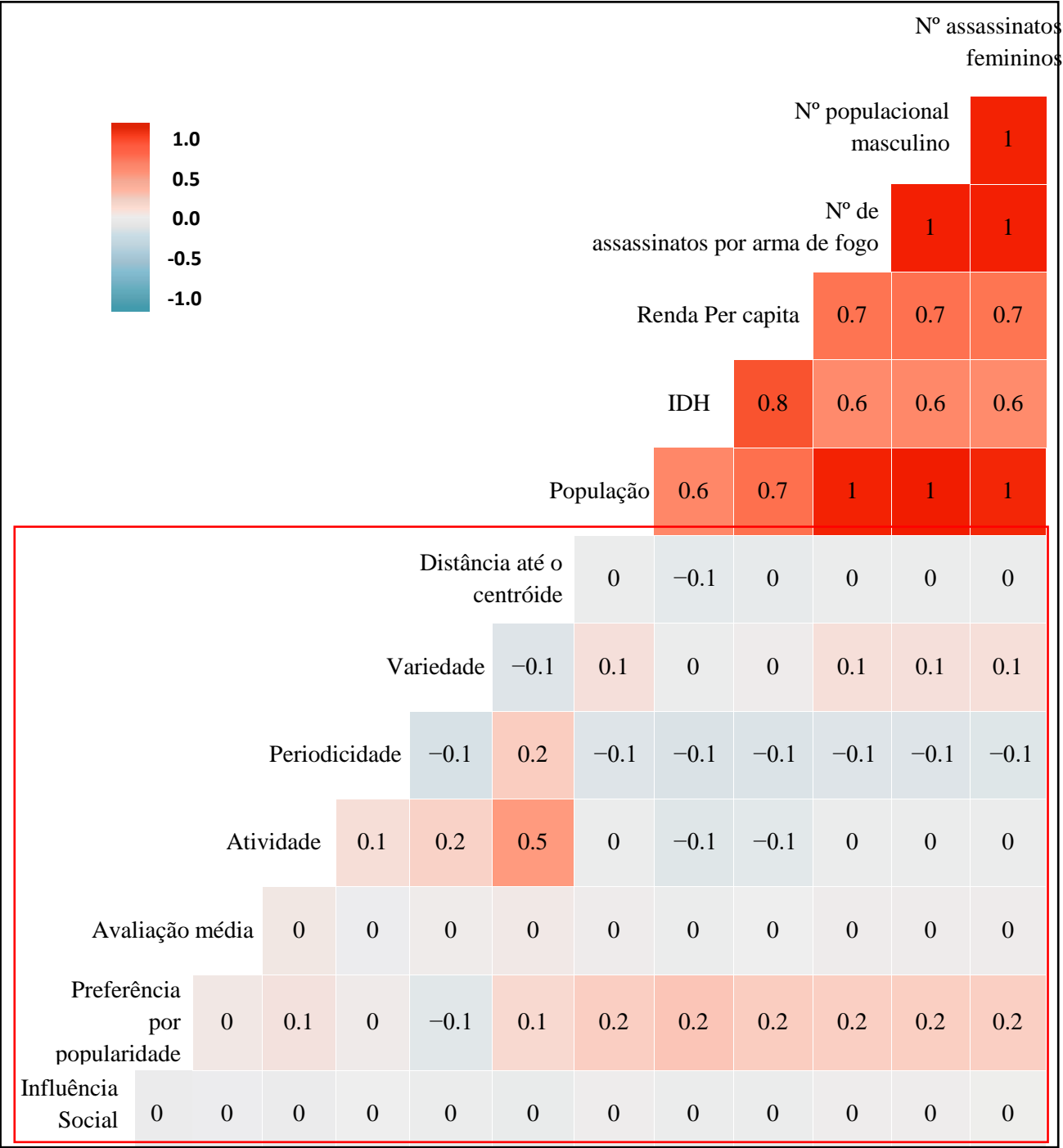
Figura 16. Correlação de Kendall dos padrões de comportamento ao contribuir dos criadores medianos e as características urbanas das cidades



Fonte: Própria Autora.

Os resultados da Figura 17 se referem aos criadores interessados em locais populares e mostram que esse tipo de colaborador não tem seu padrão de comportamento ao contribuir afetado por fatores urbanos socioeconômicos.

Figura 17. Correlação de Kendall dos padrões de comportamento ao contribuir dos criadores interessados em locais populares e as características urbanas das cidades



Fonte: Própria Autora.

Figura 18. Correlação de Kendall dos padrões de comportamento ao contribuir dos criadores produtores e as características urbanas das cidades

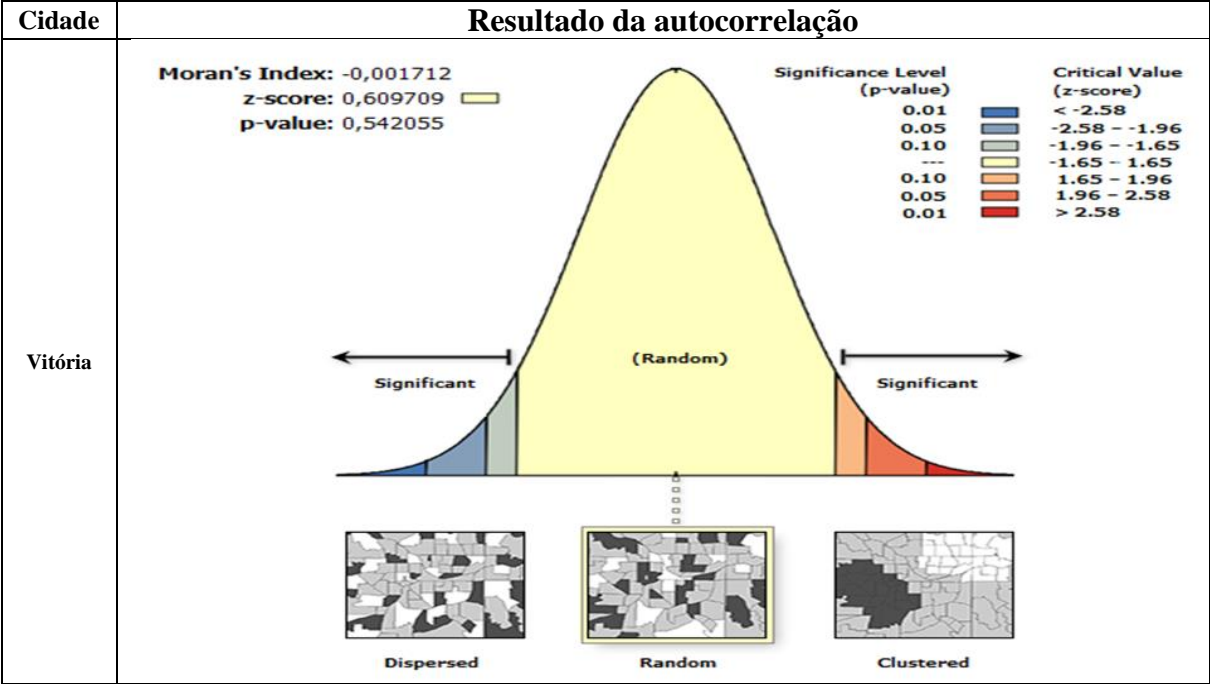


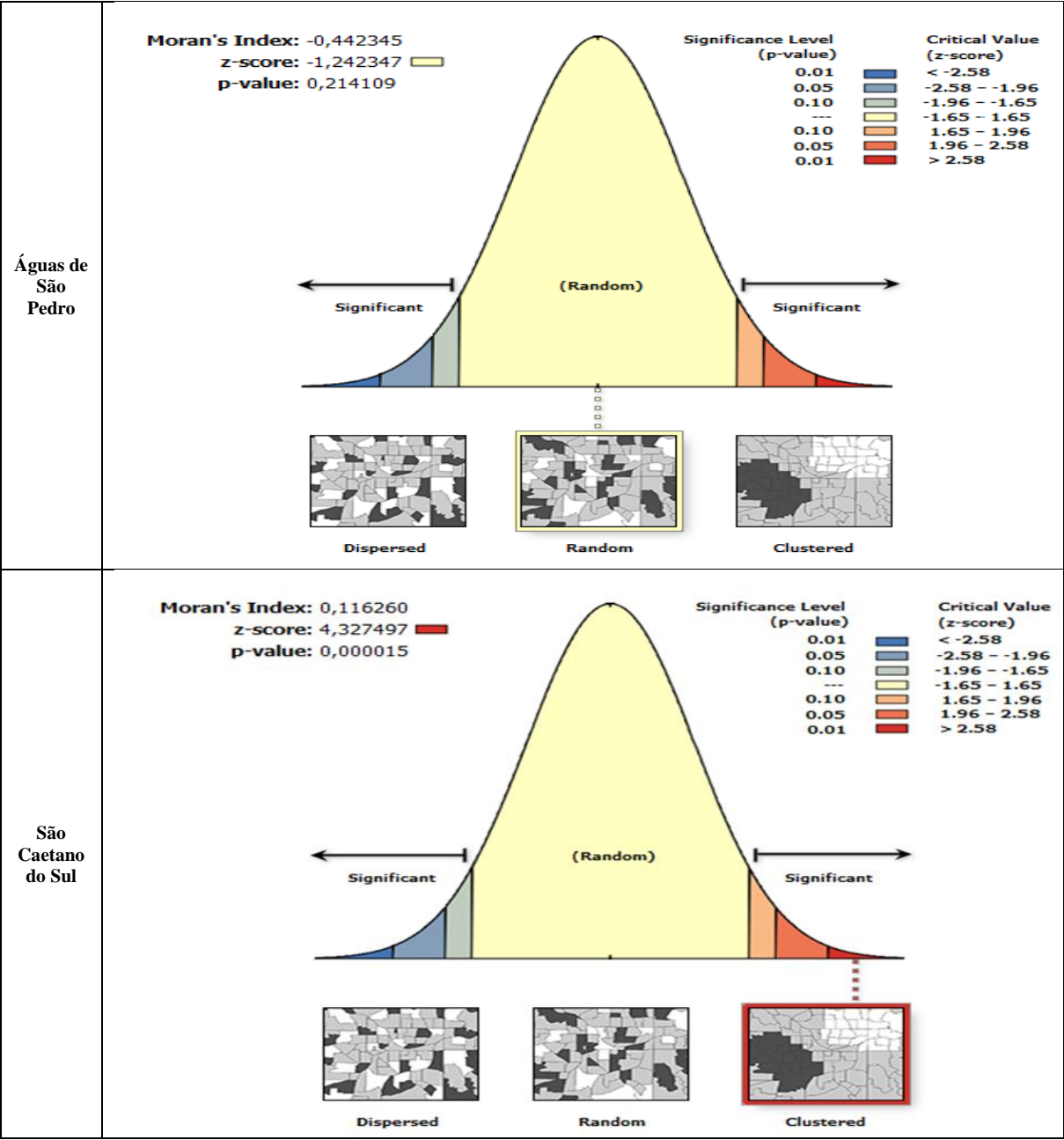
Apêndice E – Resultados da autocorrelação espacial, sob o enfoque do índice de Moran

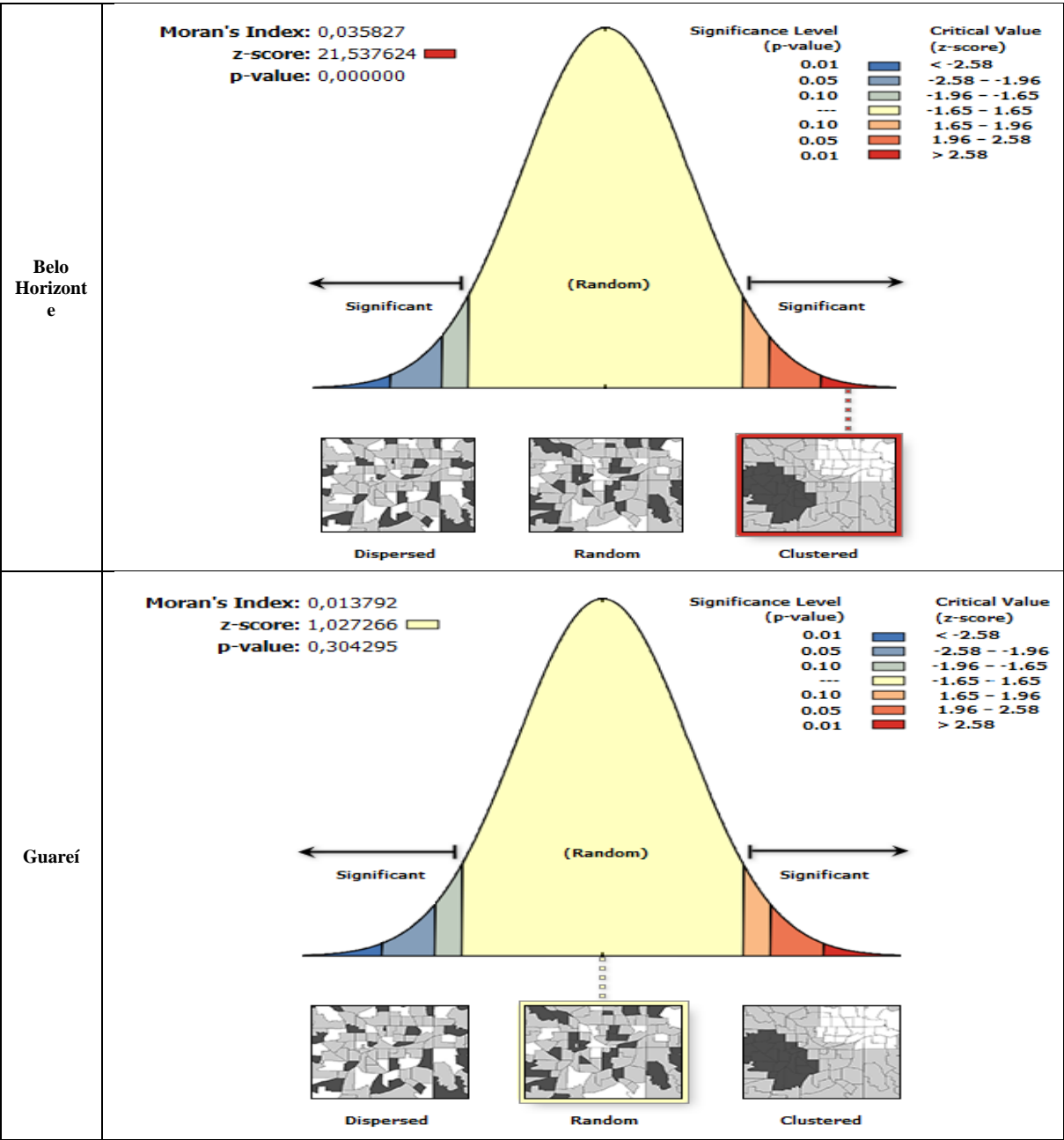
Este apêndice contém as descrições dos resultados gerados pelo método da autocorrelação espacial. Primeiramente, avaliação dos índices de Moran é detalhada nas cidades do Sudeste, na Tabela 3. Os resultados estatisticamente válidos são aqueles que apresentam p-valor menor do que 0,005. Quem não apresentar tal comportamento, não pode afirmar nada sobre o comportamento espacial dos criadores da cidade. Isso acontece nas cidades de Vitória, Águas de São Pedro, Guareí e Santos.

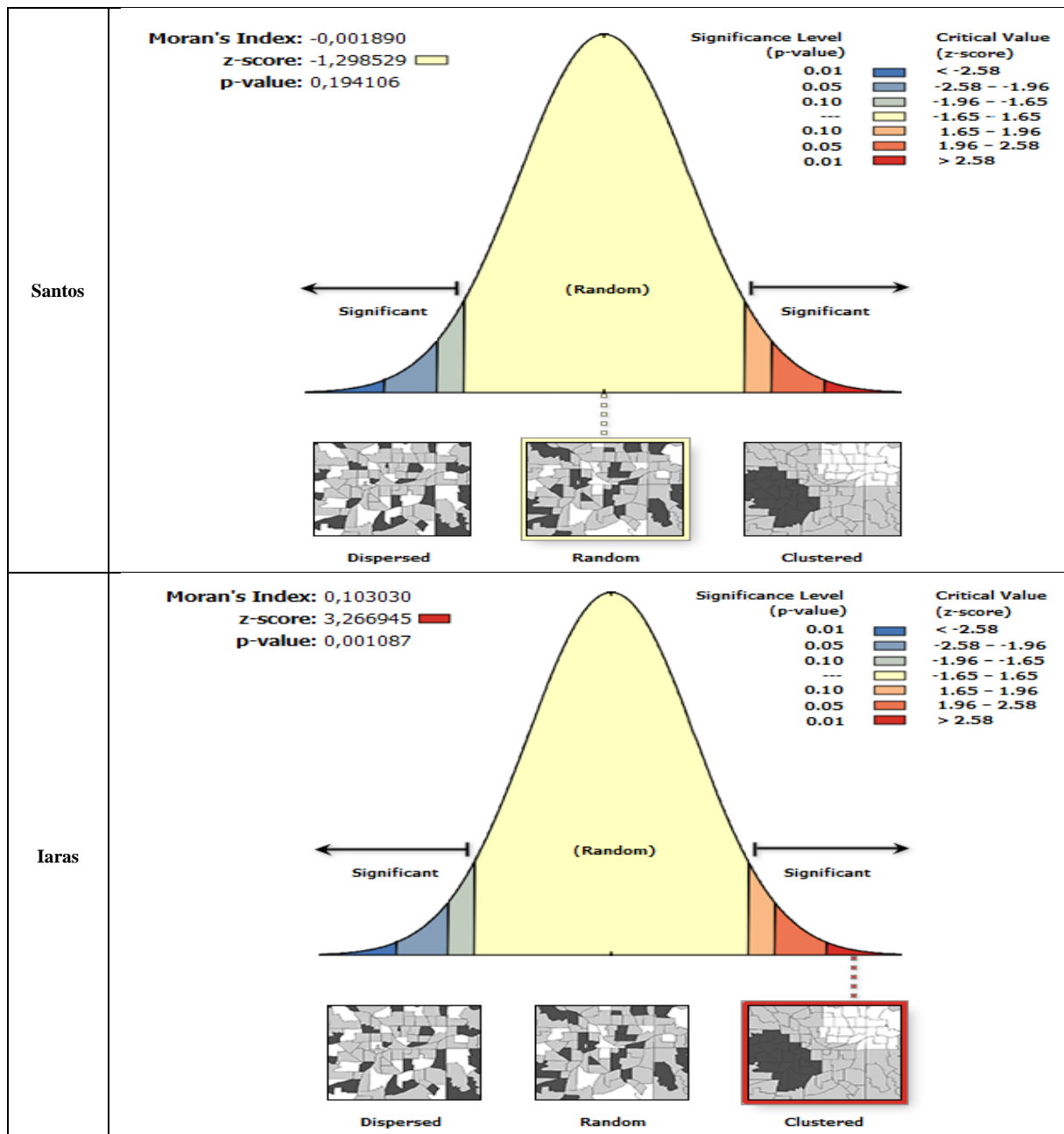
As cidades de São Caetano do Sul, Belo Horizonte, Iaras, São Paulo e Rio de Janeiro apresentaram um comportamento espacial clusterizado. Isso significa que eles colaboram com a RG em setores censitários próximos.

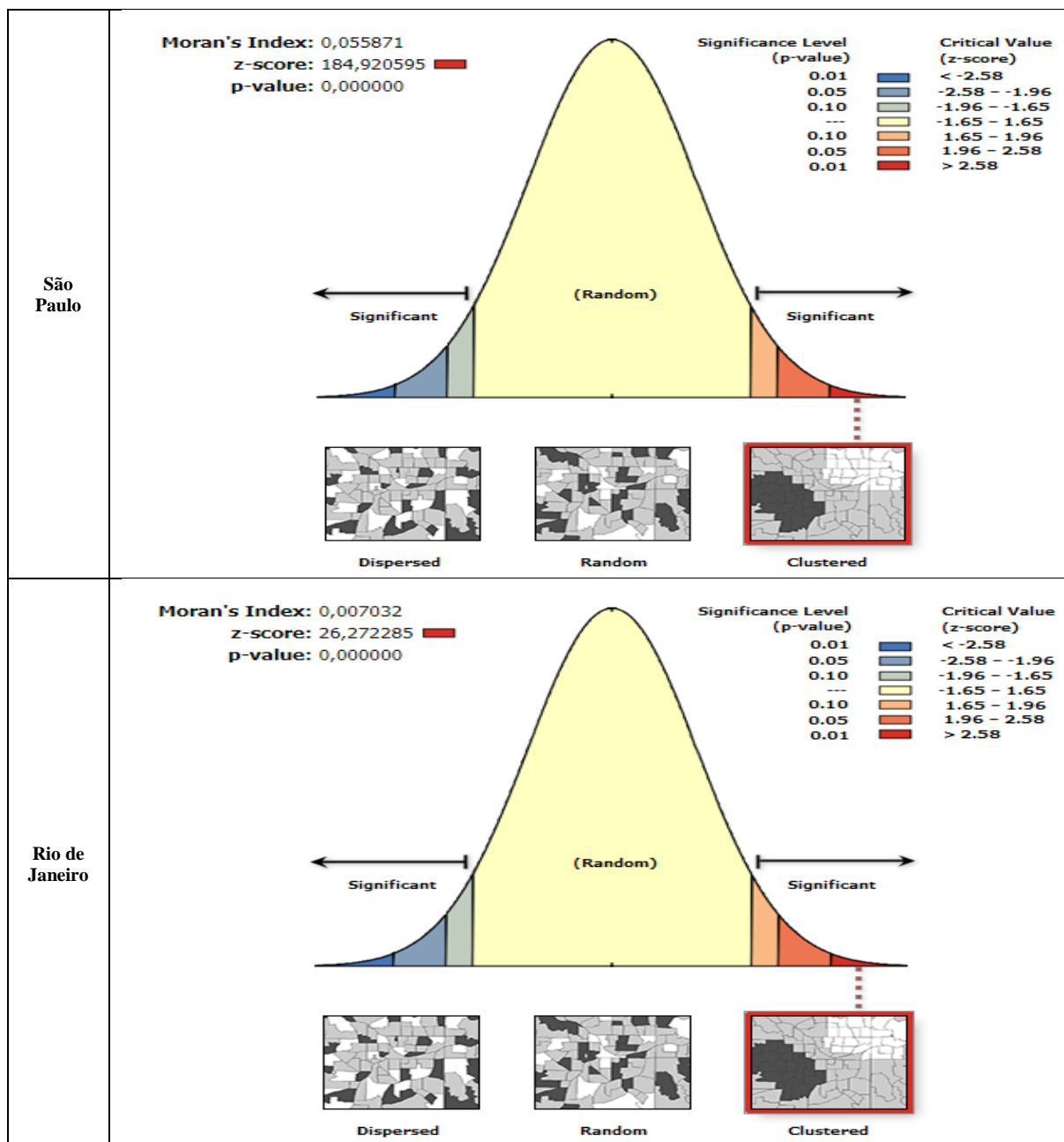
Tabela 3. Resultados da autocorrelação espacial em cidades do Sudeste







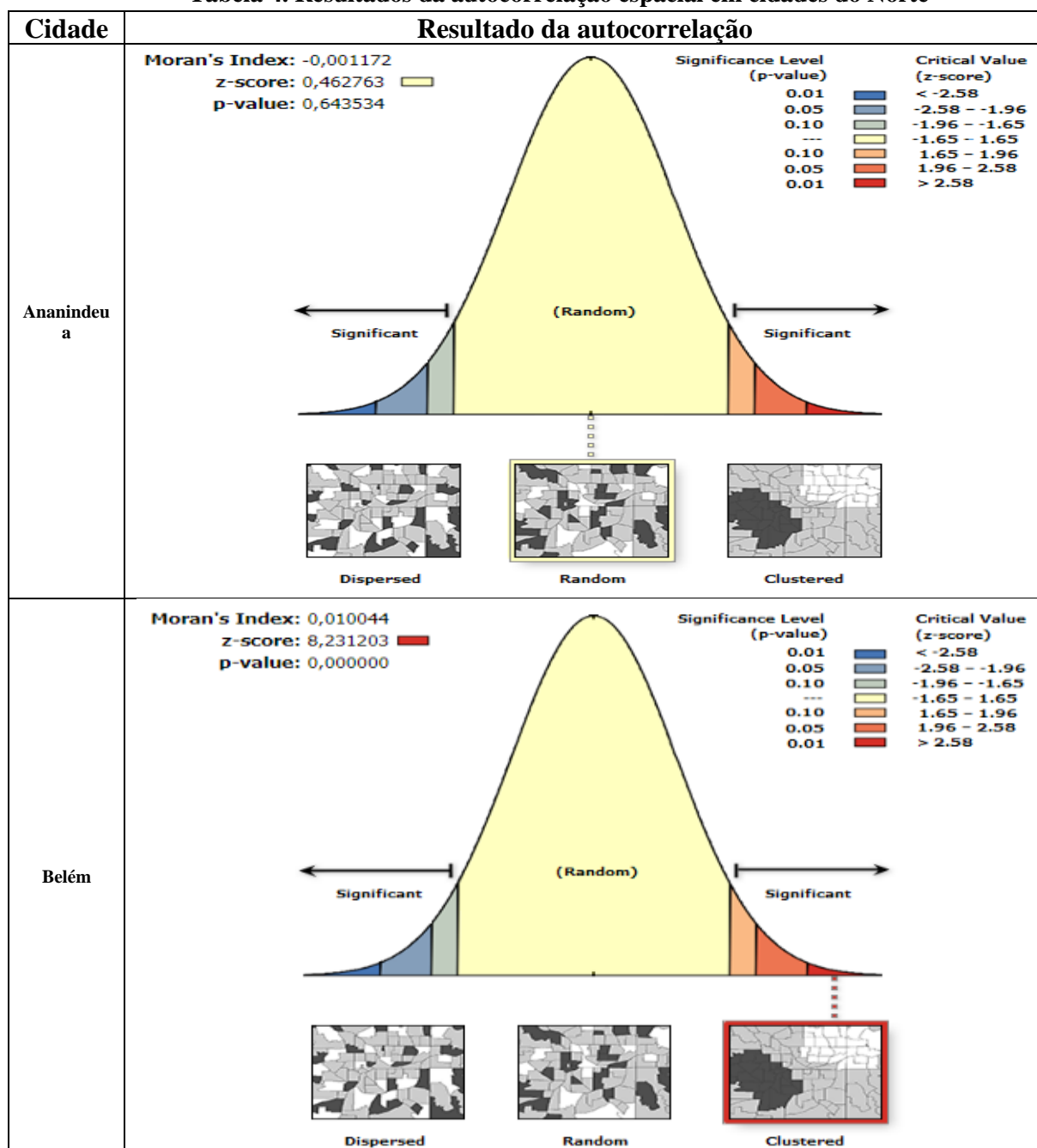


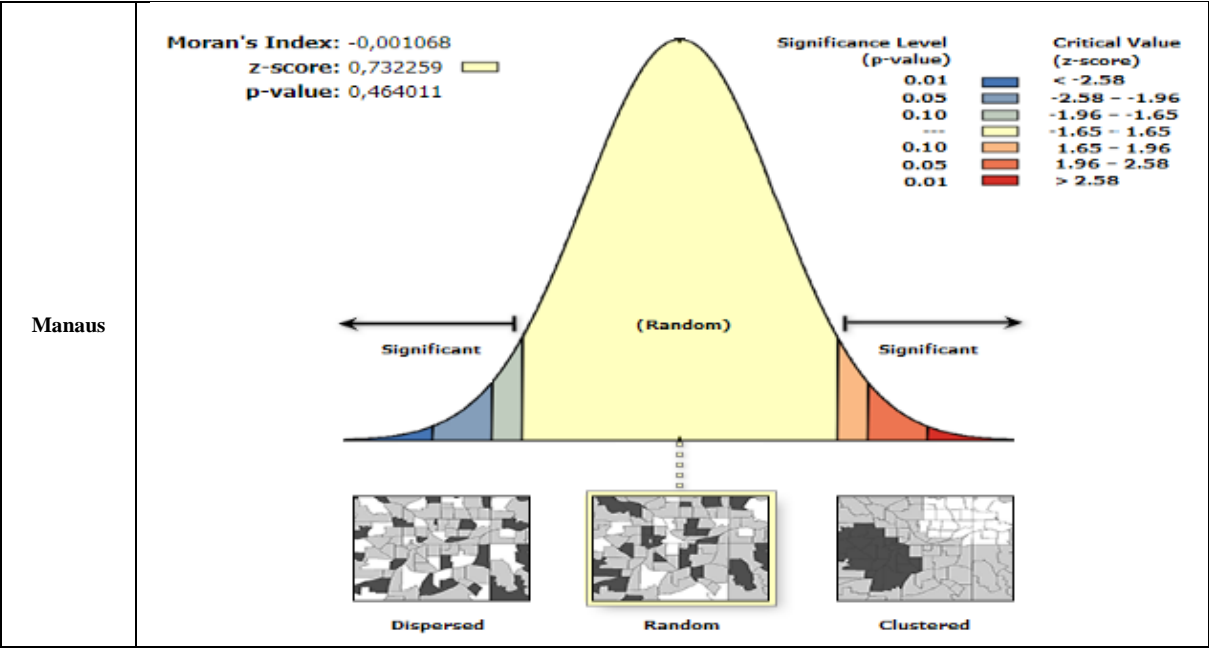


Fonte: Própria Autora.

A Tabela 4 detalha o comportamento espacial da região Norte, de acordo com o índice de Moran. Os resultados mostram que somente a cidade de Belém apresenta um comportamento espacial agrupado entre as cidades observadas da região Norte. Por outro lado, as cidades de Ananindeua e Manaus apresentam um p-valor maior que 0,005 e, por isso, não há como afirmar com validade estatística sobre o comportamento espacial dessas cidades.

Tabela 4. Resultados da autocorrelação espacial em cidades do Norte

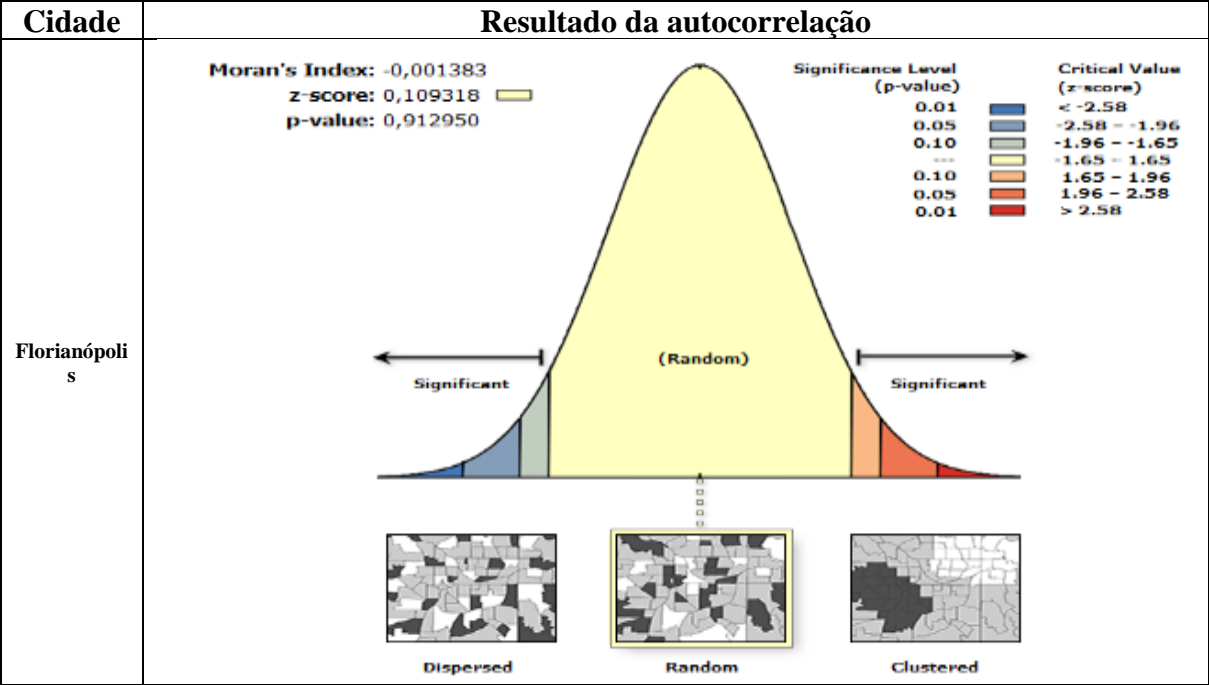


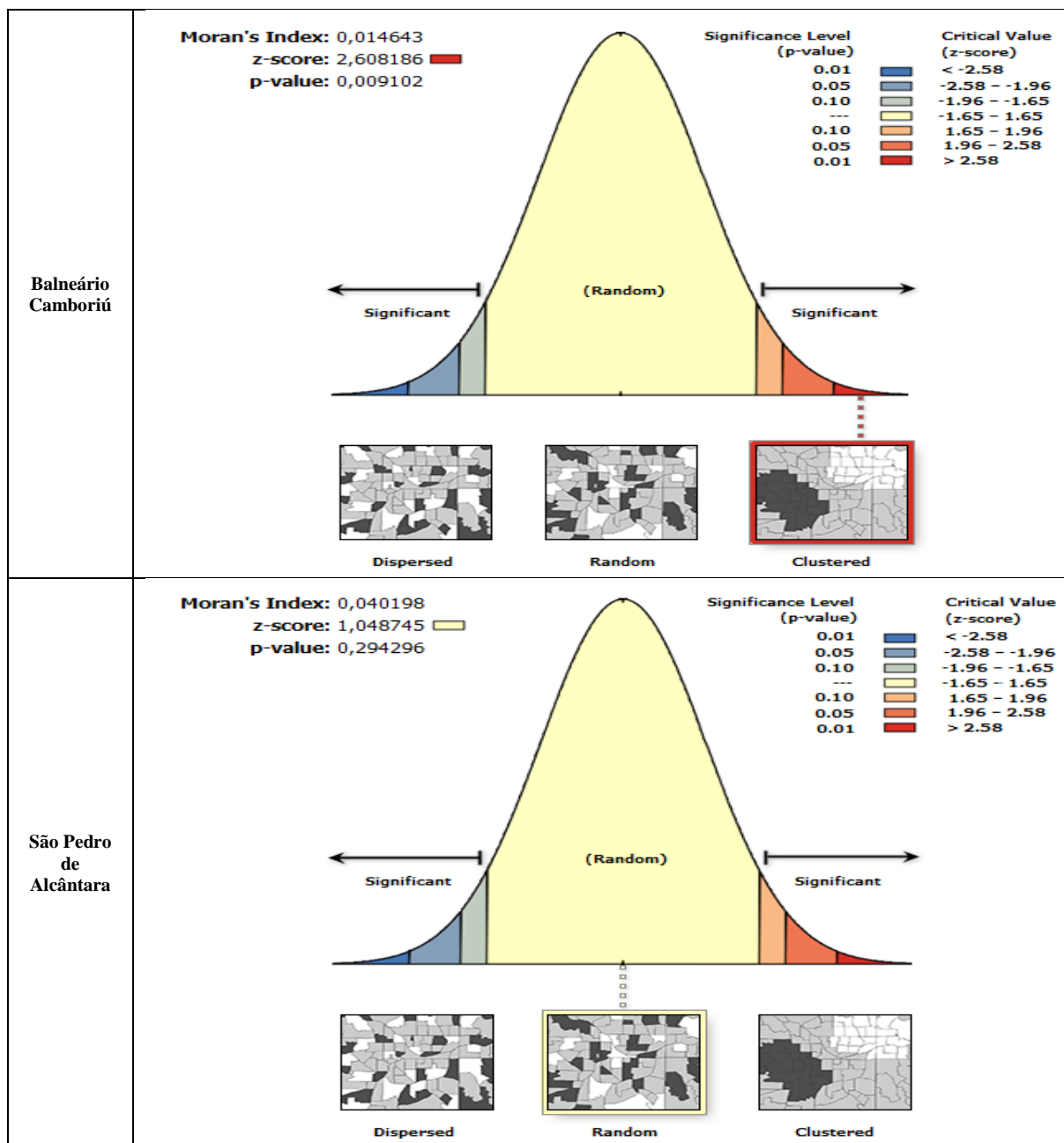


Fonte: Própria Autora.

A Tabela 5 detalha os resultados da autocorrelação das cidades da região Sul do Brasil. De acordo com os resultados, apenas a cidade de Balneário Camboriú possui um padrão espacial agrupado sobre as participações feitas nos setores censitários.

Tabela 5. Resultados da autocorrelação espacial em cidades do Sul

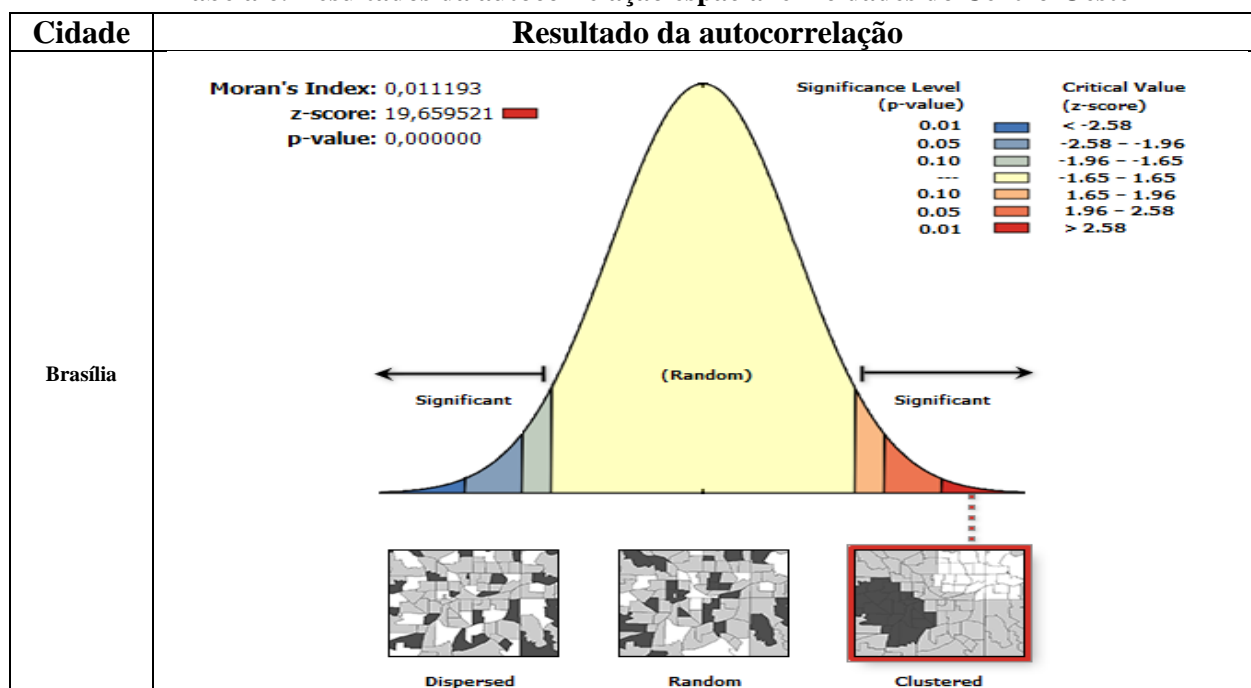




Fonte: Própria Autora.

A única cidade selecionada da região Centro-Oeste é Brasília. Os resultados do índice de Moran indicam que a capital do país possui participantes de RG que se agrupam em setores censitários próximos.

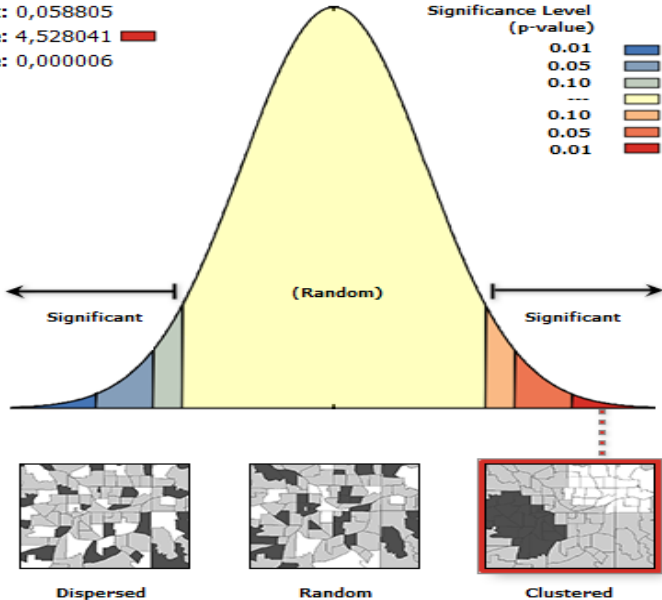
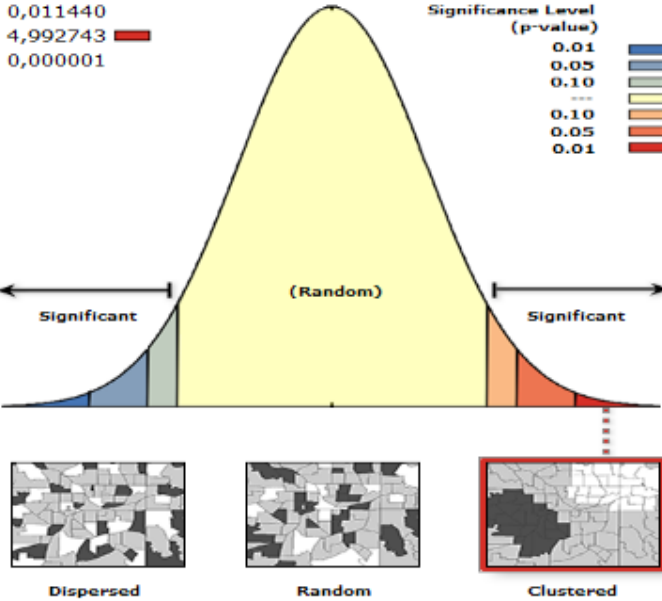
Tabela 6. Resultados da autocorrelação espacial em cidades do Centro-Oeste

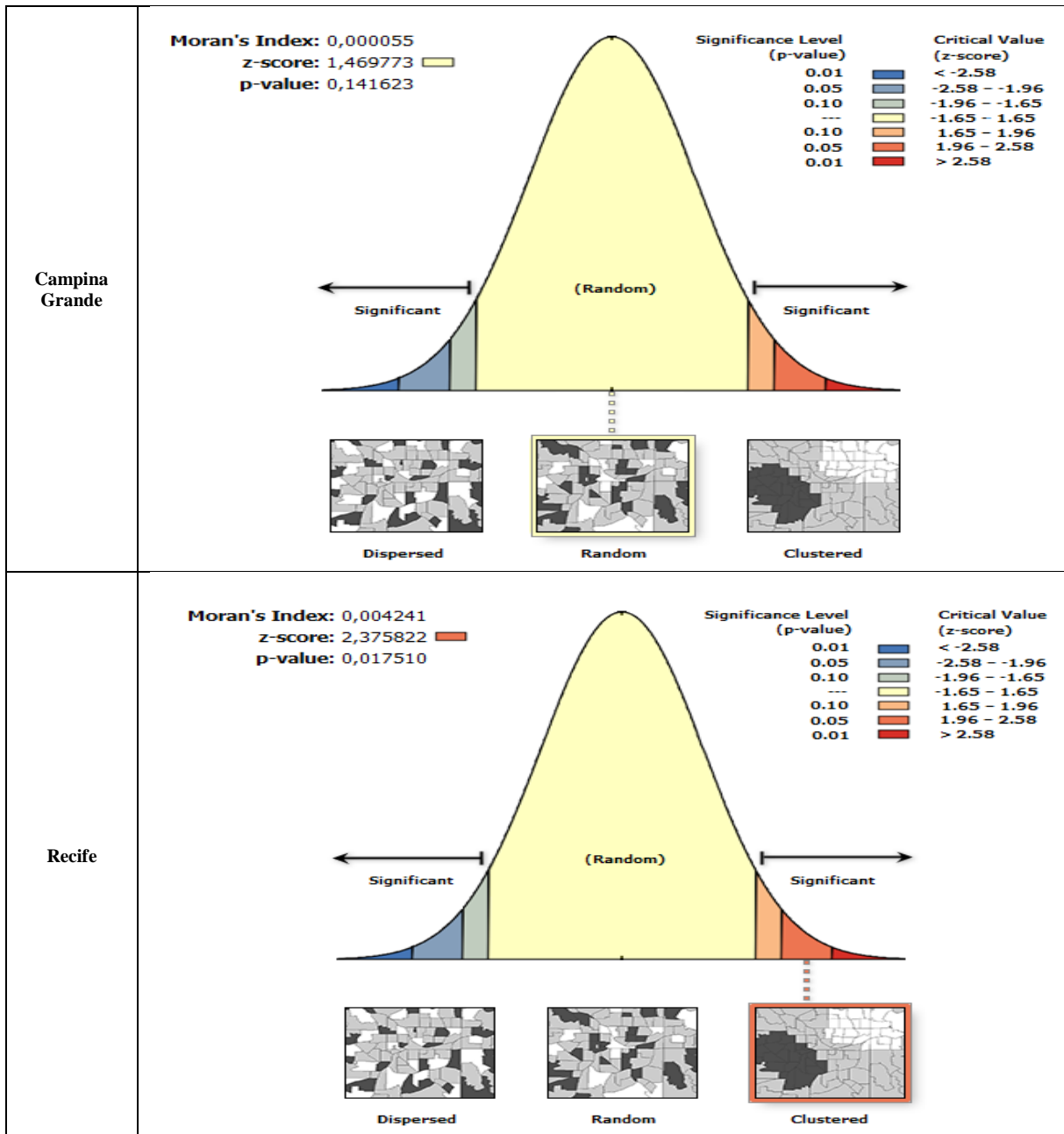


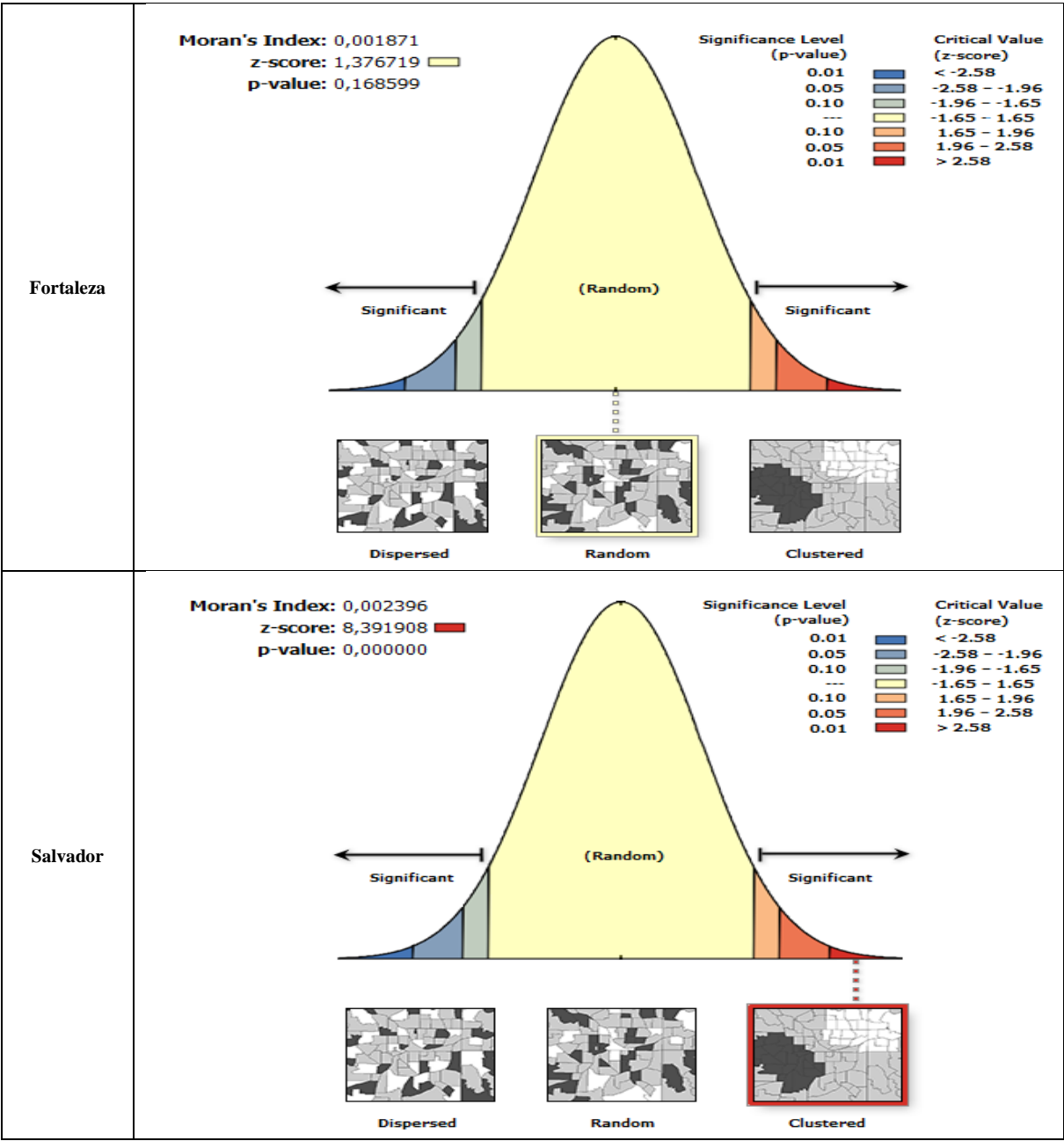
Fonte: Própria Autora.

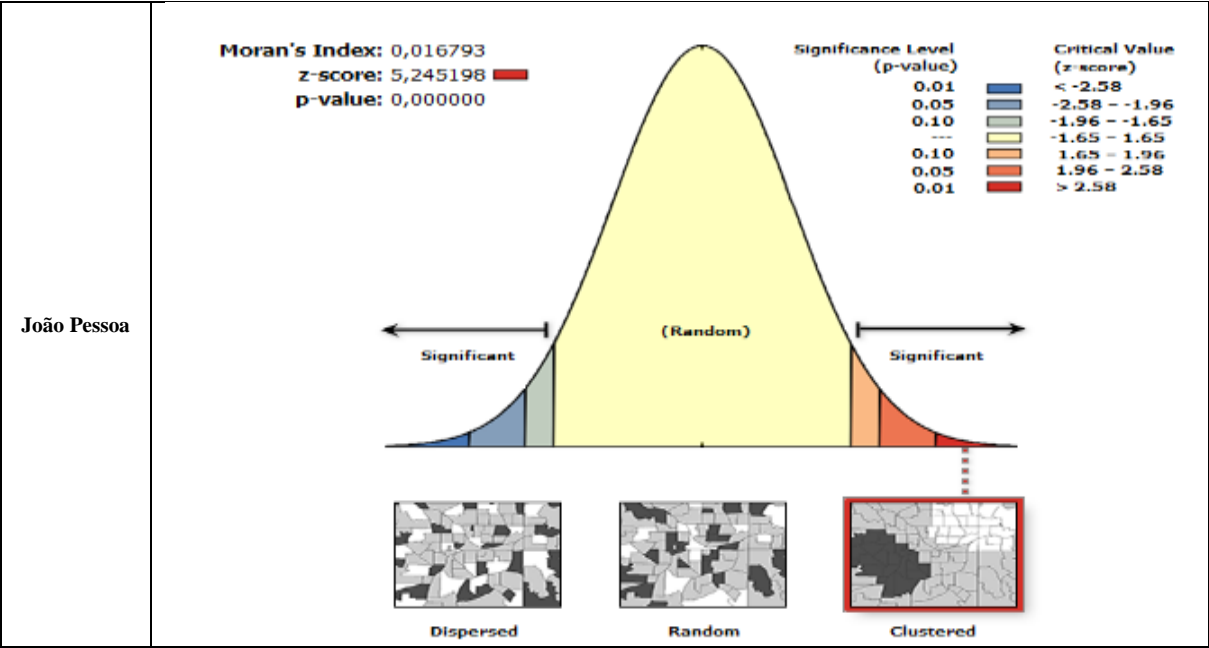
A última região avaliada é a região Nordeste. O índice de Moran detecta um comportamento agrupado da participação dos seus colaboradores, em setores das cidades de Olinda, Caruaru, Recife, Salvador e João Pessoa.

Tabela 7. Resultados da autocorrelação espacial em cidades do Nordeste

Cidade	Resultado da autocorrelação
Olinda	<p> Moran's Index: 0,058805 z-score: 4,528041 p-value: 0,000006 </p>  <p> Significance Level (p-value) 0.01 0.05 0.10 --- 0.10 0.05 0.01 </p> <p> Critical Value (z-score) < -2.58 -2.58 - -1.96 -1.96 - -1.65 --- 1.65 - 1.96 1.96 - 2.58 > 2.58 </p> <p> (Random) Significant Significant Dispersed Random Clustered </p>
Caruaru	<p> Moran's Index: 0,011440 z-score: 4,992743 p-value: 0,000001 </p>  <p> Significance Level (p-value) 0.01 0.05 0.10 --- 0.10 0.05 0.01 </p> <p> Critical Value (z-score) < -2.58 -2.58 - -1.96 -1.96 - -1.65 --- 1.65 - 1.96 1.96 - 2.58 > 2.58 </p> <p> (Random) Significant Significant Dispersed Random Clustered </p>





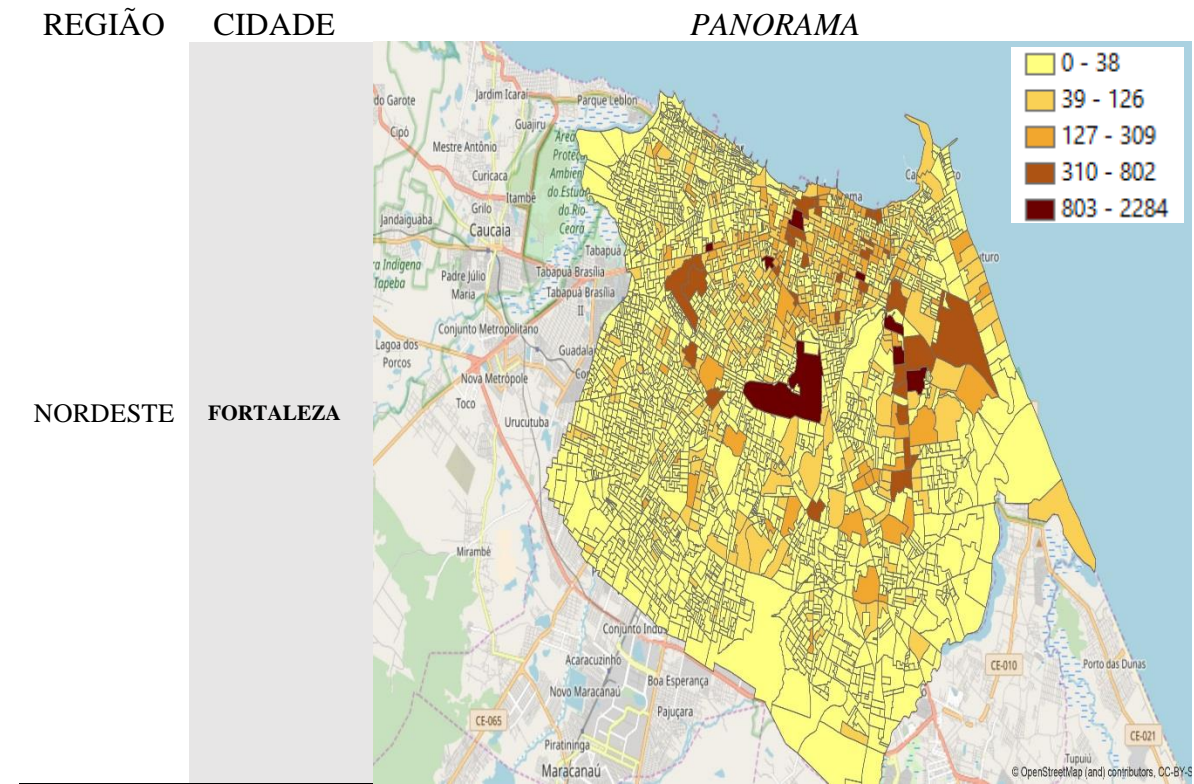


Fonte: Própria Autora.

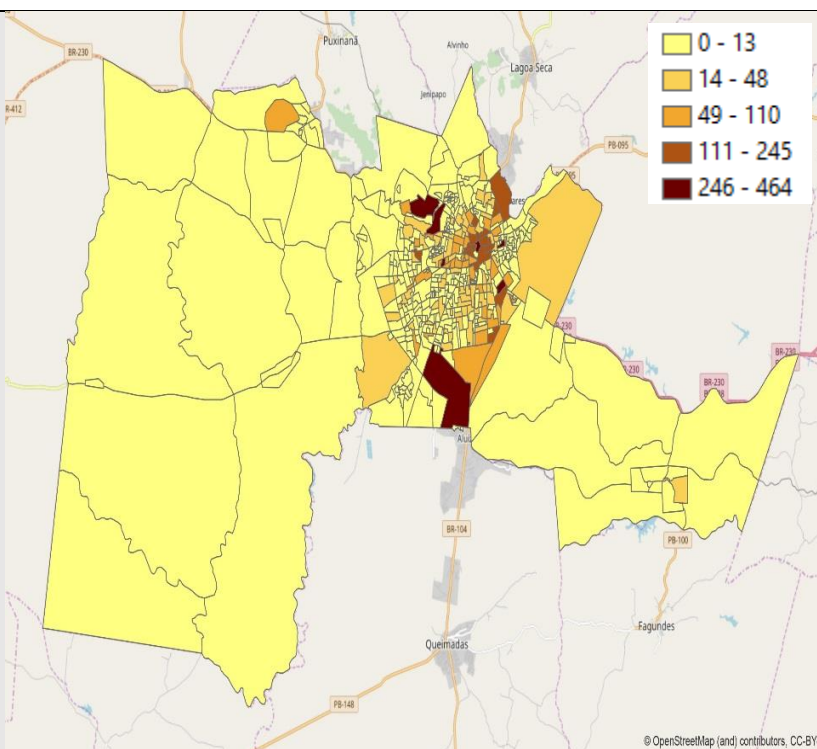
Apêndice F – Distribuição da participação de criadores de conteúdo geolocalizado em cada cidade brasileira selecionada

Neste apêndice, é mostrado a participação quantitativa dos criadores de conteúdo, nos setores censitários. O objetivo deste detalhamento é ter um panorama geral das atividades nos setores censitários das cidades selecionadas para comparar com os resultados das estatísticas espaciais presentes nesta tese. Os mapas podem ser vistos a seguir no Quadro 7.

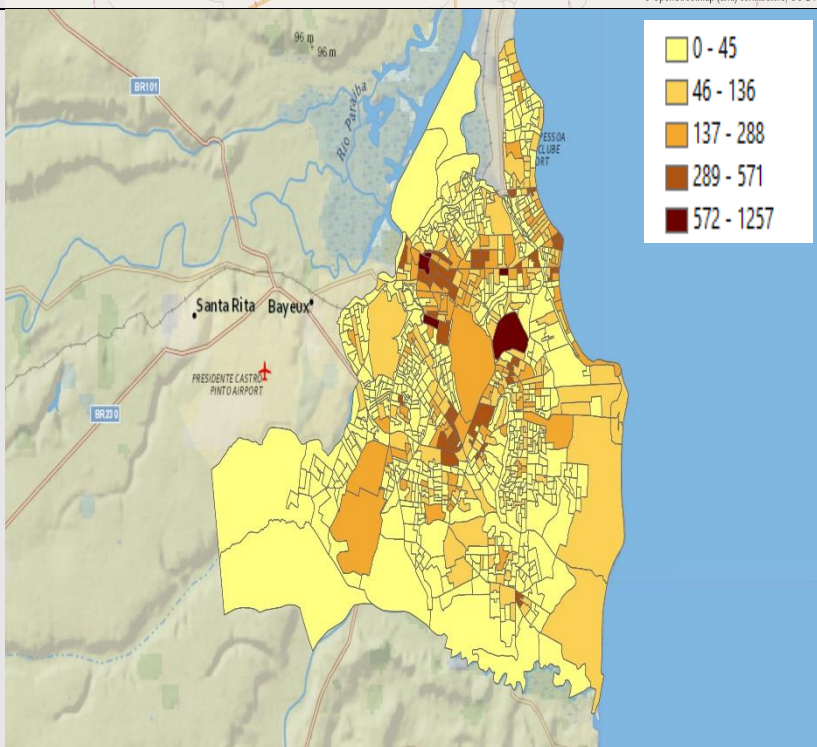
Quadro 7. Distribuição da participação dos colaboradores nas áreas urbanas de cada cidade brasileira selecionada



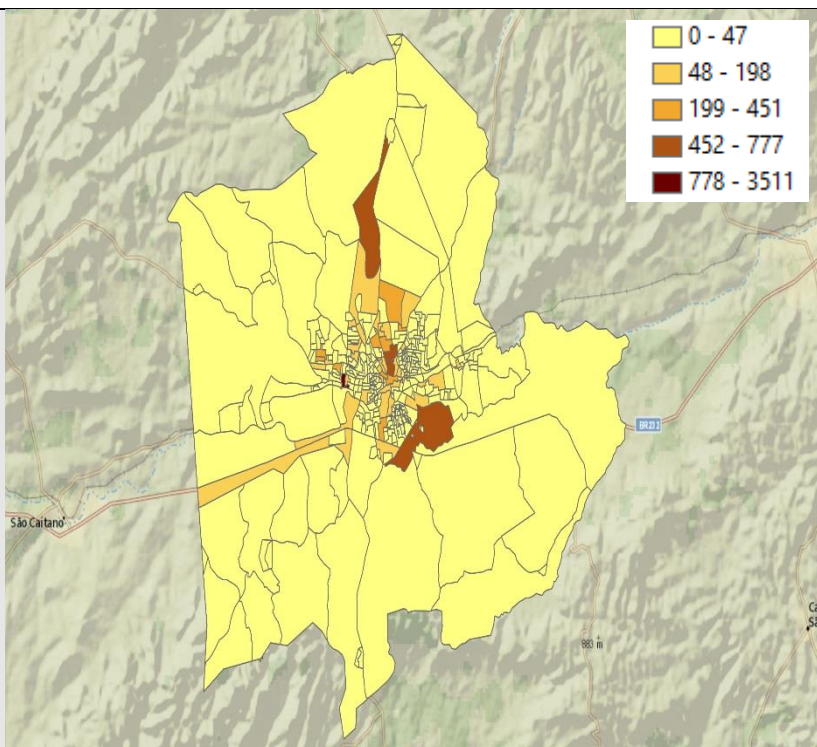
CAMPINA GRANDE



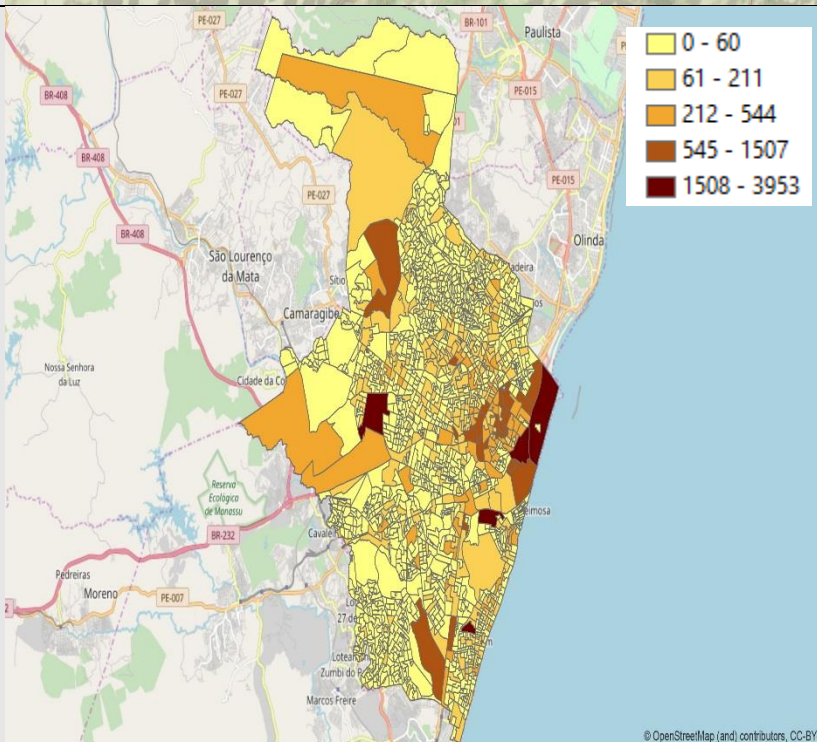
JOÃO PESSOA

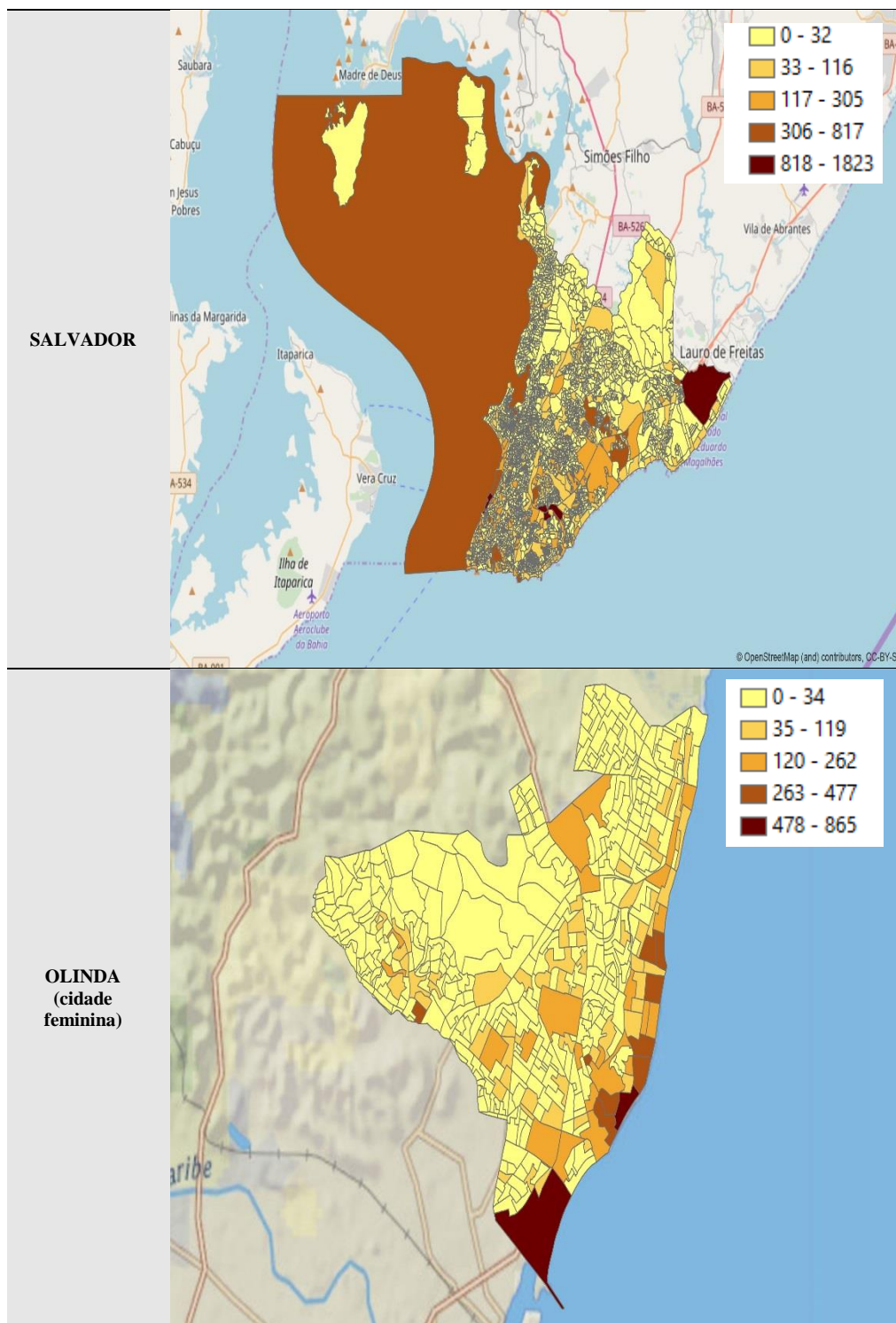


CARUARU

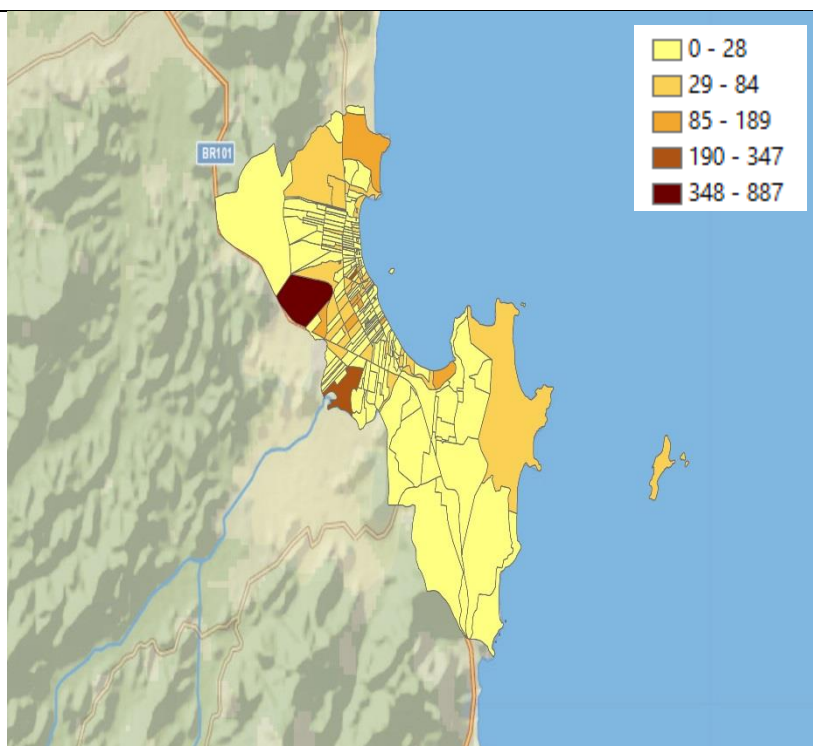


RECIFE (cidade feminina)



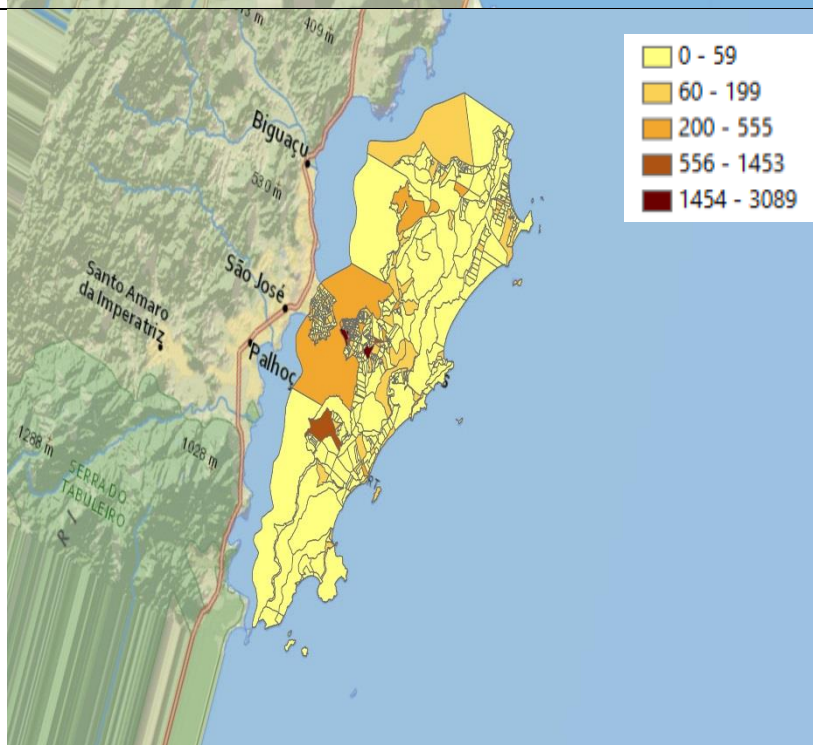


**BALNEÁRIO
CAMBORIÚ**
(5º melhor IDH
do Brasil)

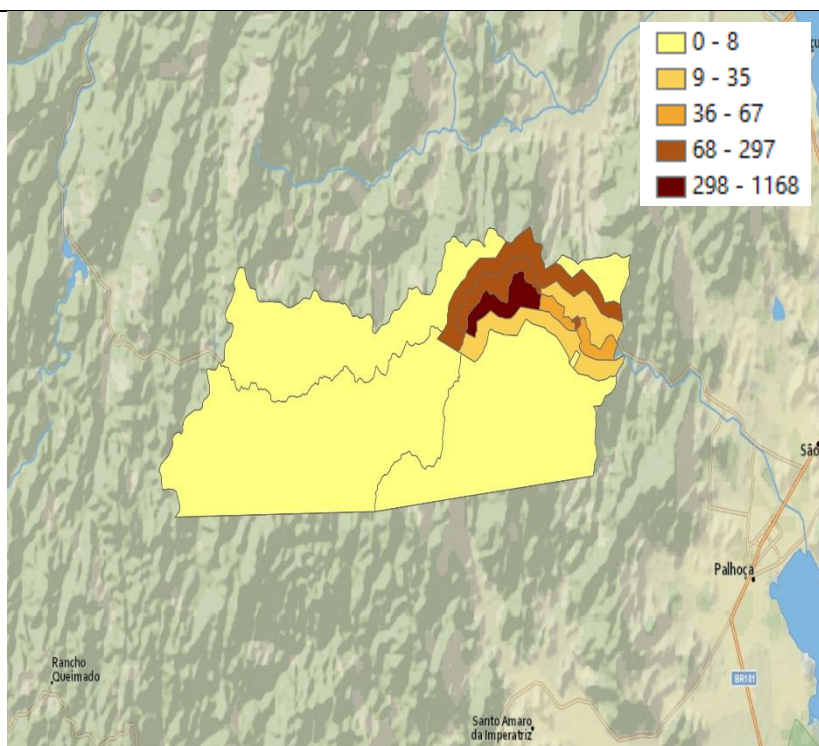


SUL

FLORIANÓPOLIS
(3º melhor IDH
do Brasil)

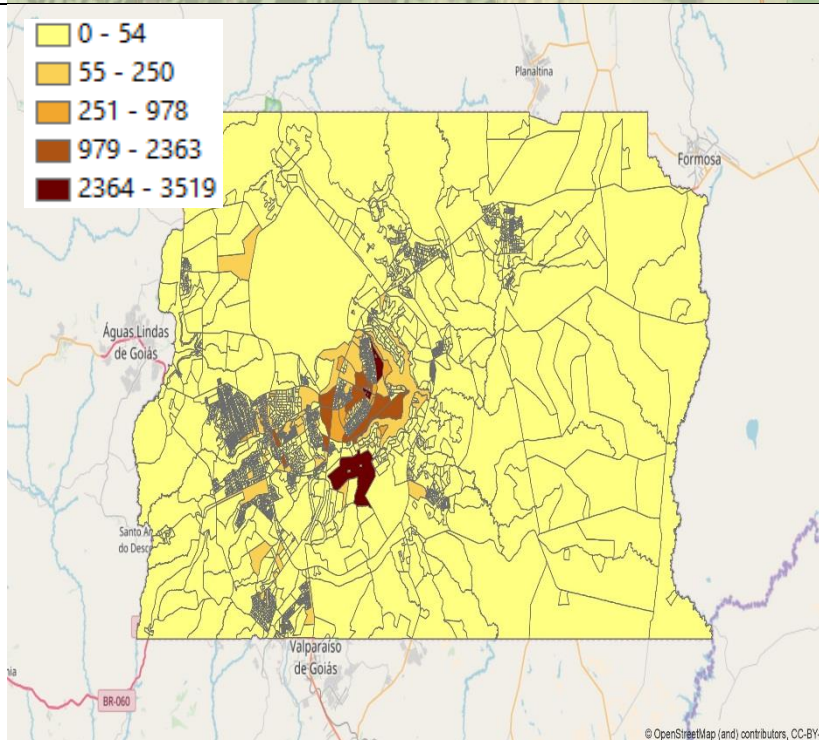


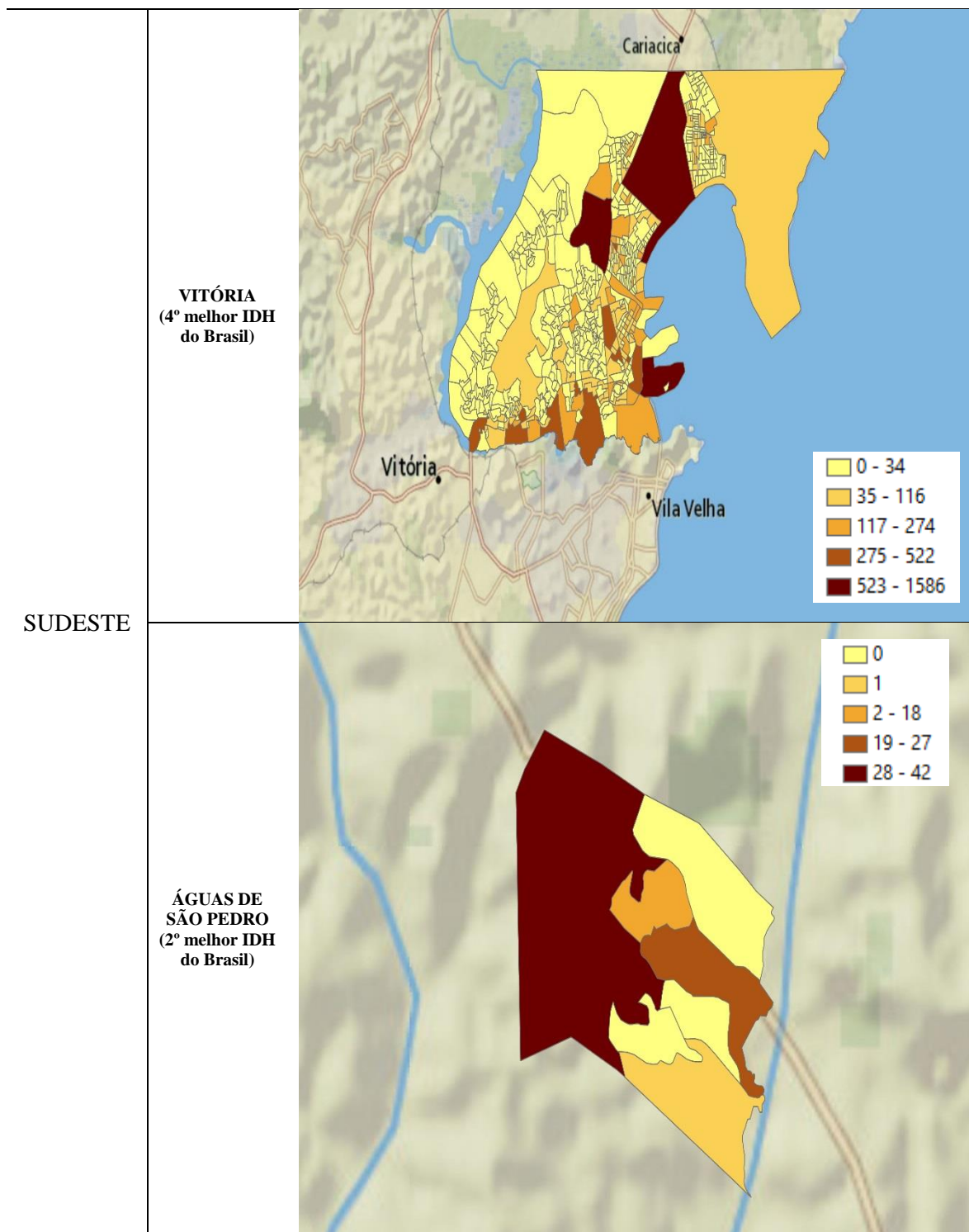
**SÃO PEDRO
DE
ALCÂNTARA**
(cidade
masculina)



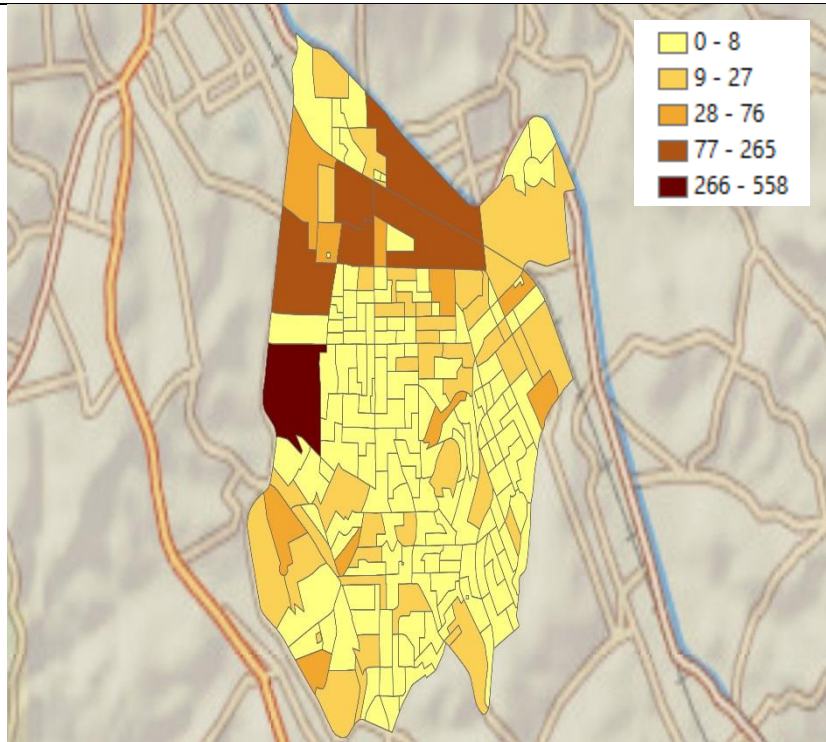
**CENTRO-
OESTE**

BRASÍLIA

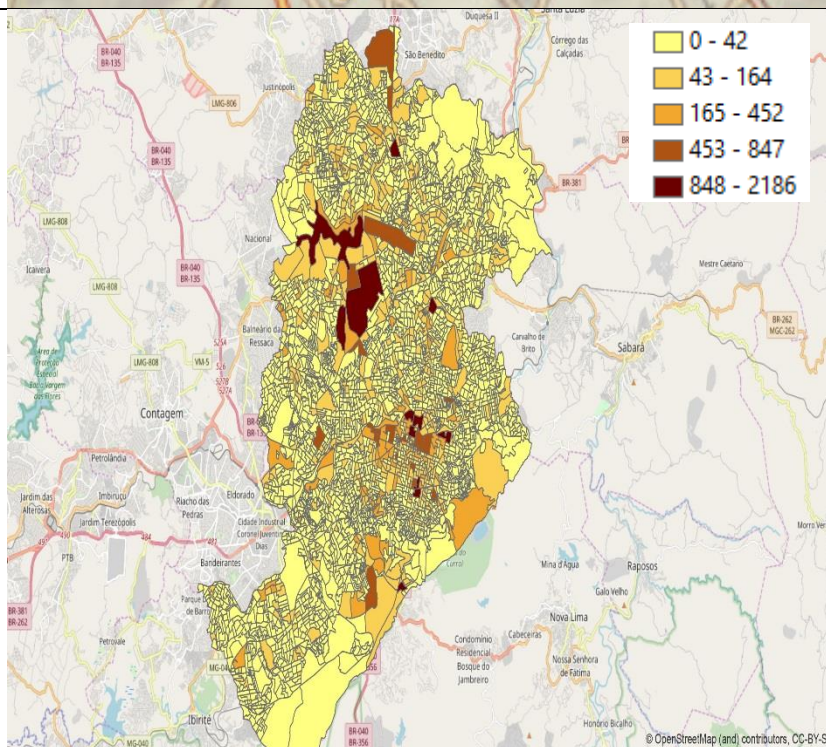




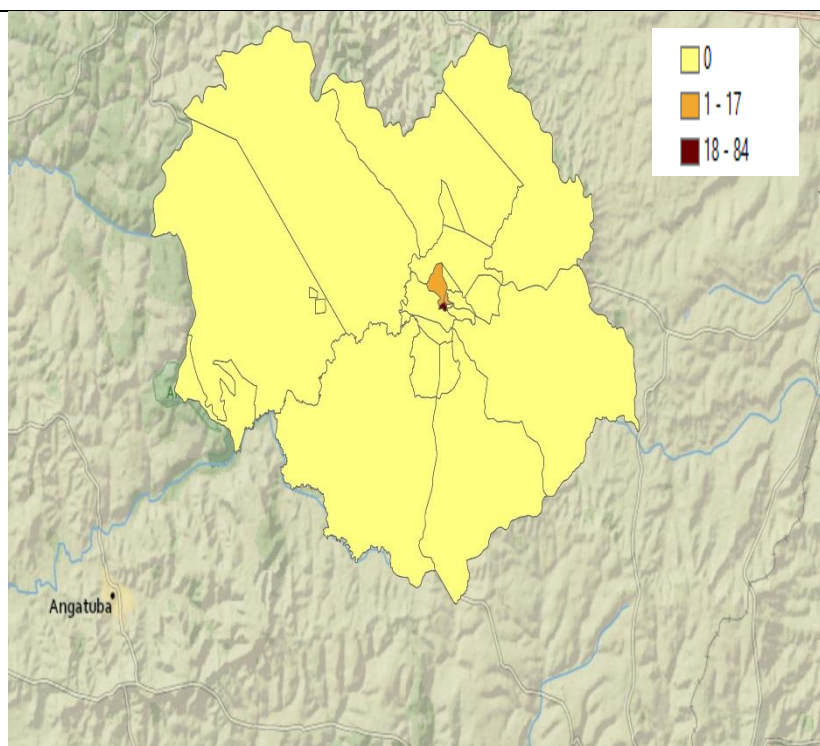
**SÃO
CAETANO
DO SUL**
(1º melhor IDH
do Brasil,
cidade
feminina)



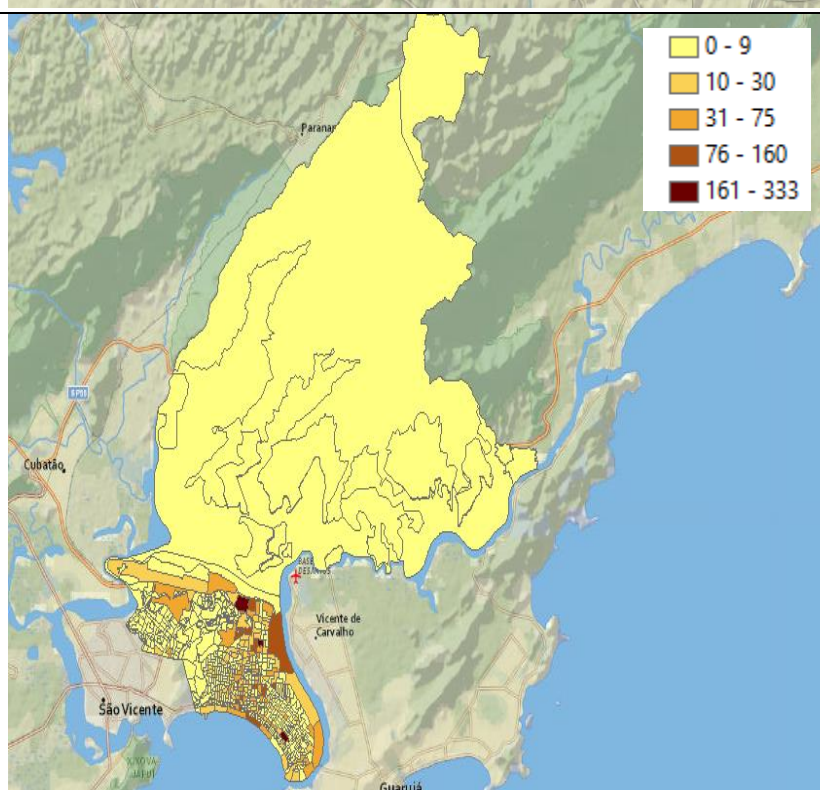
**BELO
HORIZONTE**



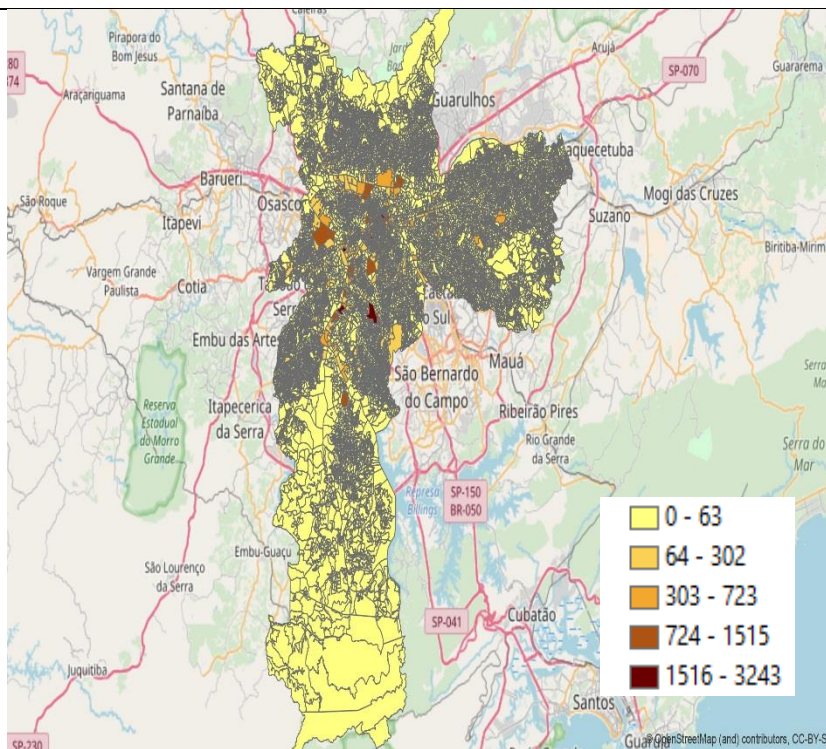
GUAREÍ
(cidade masculina)



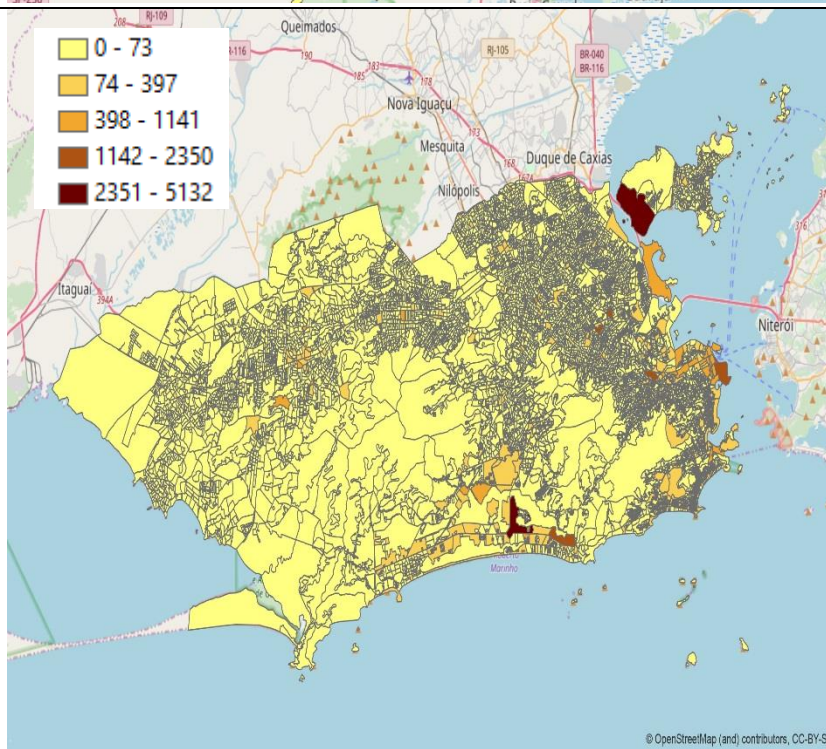
SANTOS
(cidade feminina)



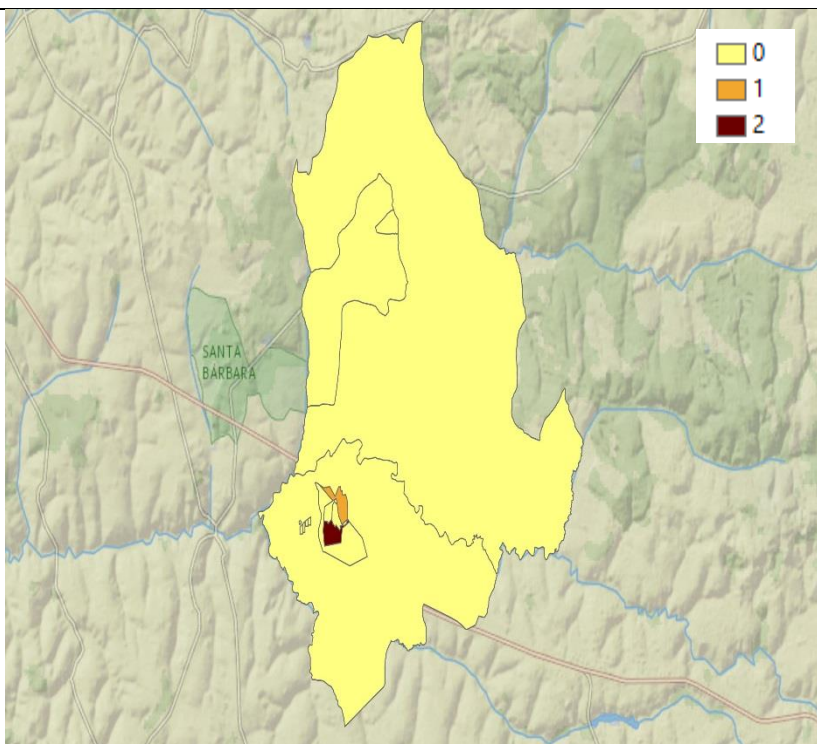
SÃO PAULO



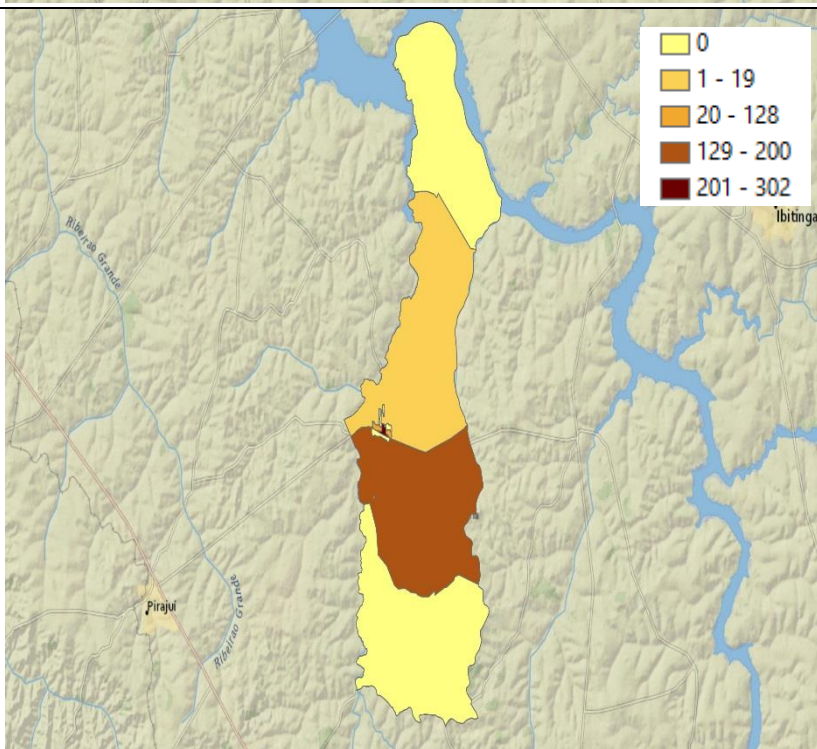
RIO DE JANEIRO



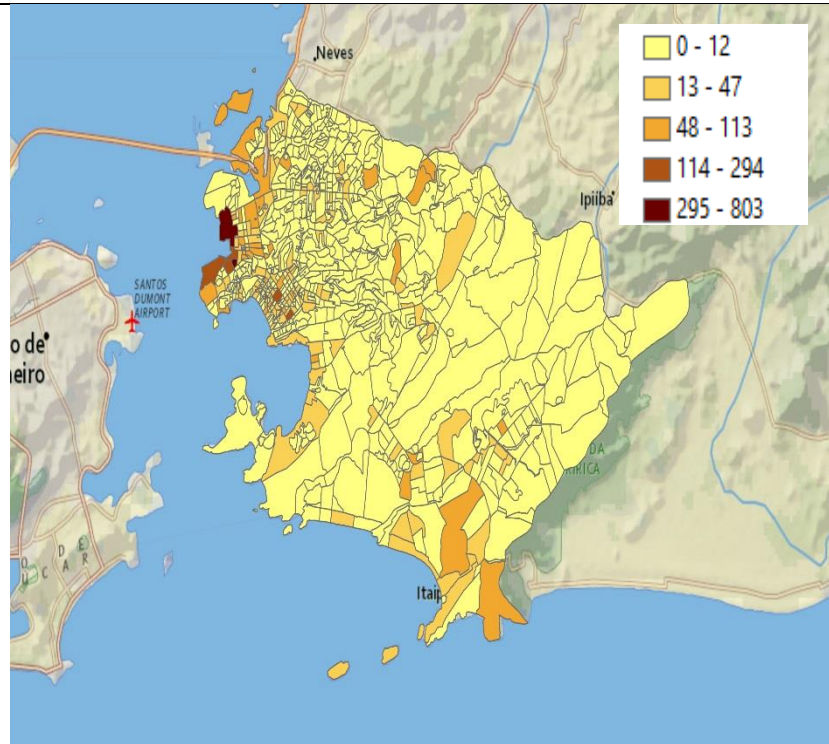
IARAS
(cidade masculina)



REGINÓPOLIS
(cidade masculina)

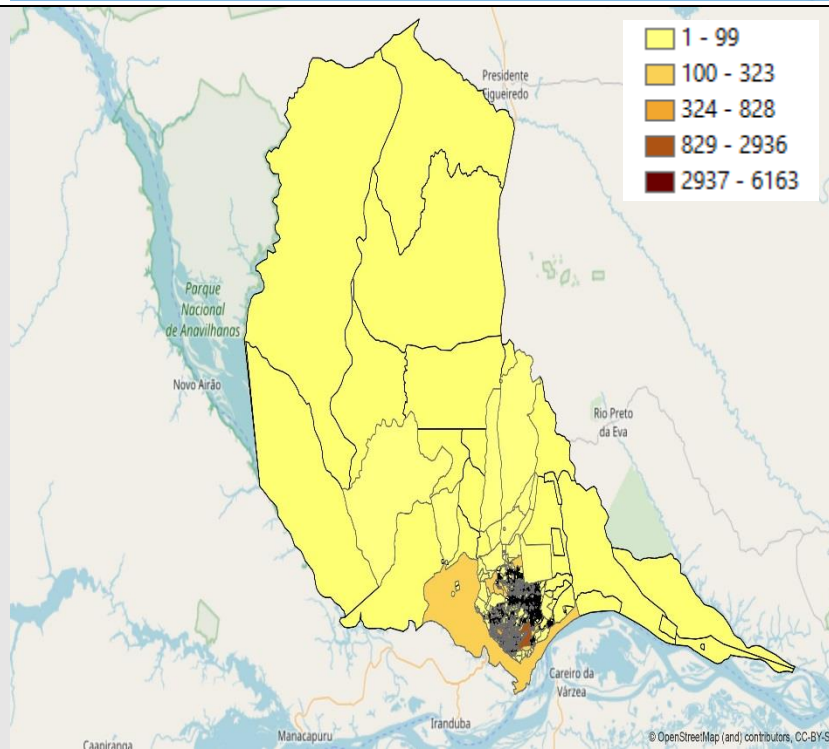


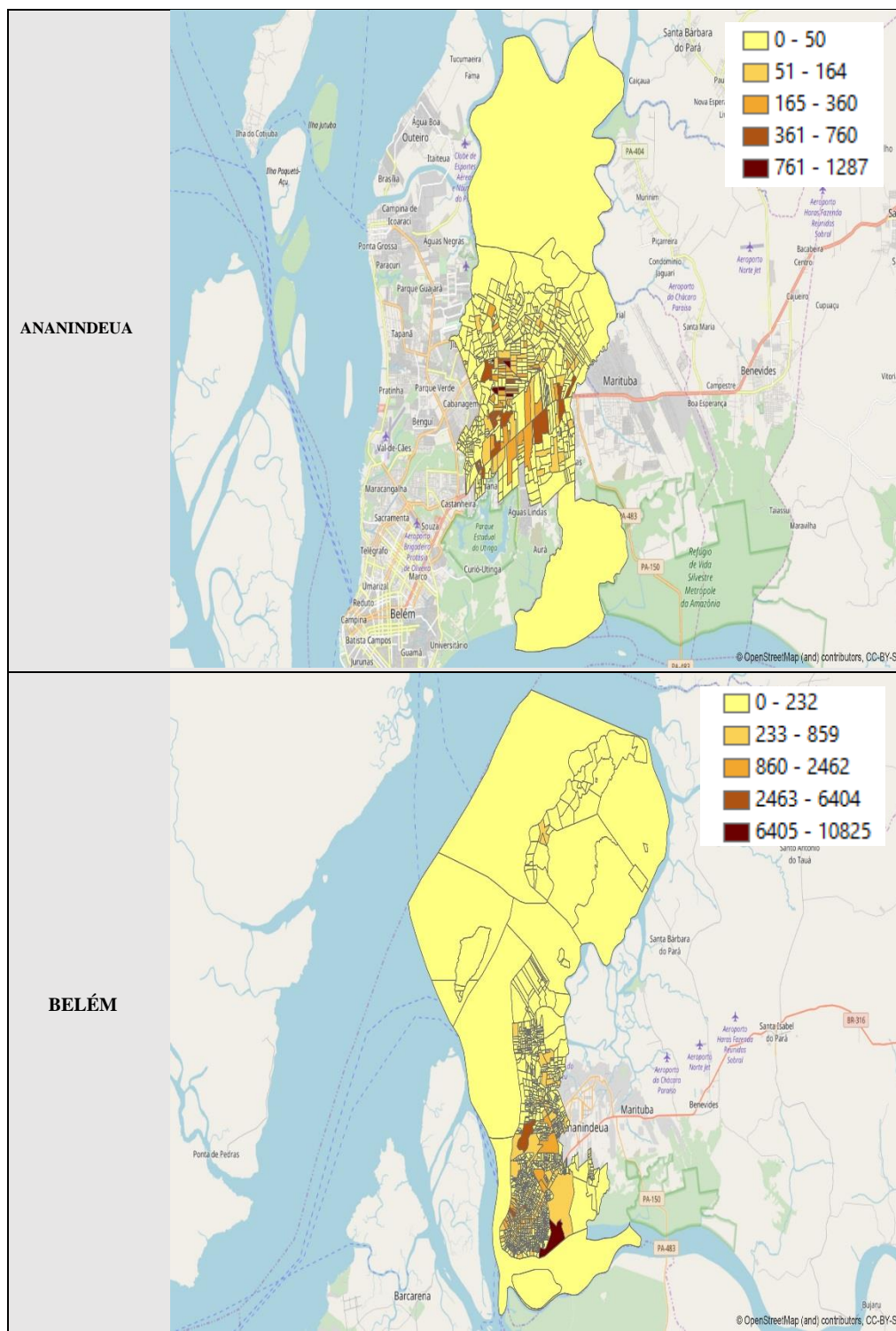
NITERÓI
(cidade
feminina)



NORTE

MANAUS



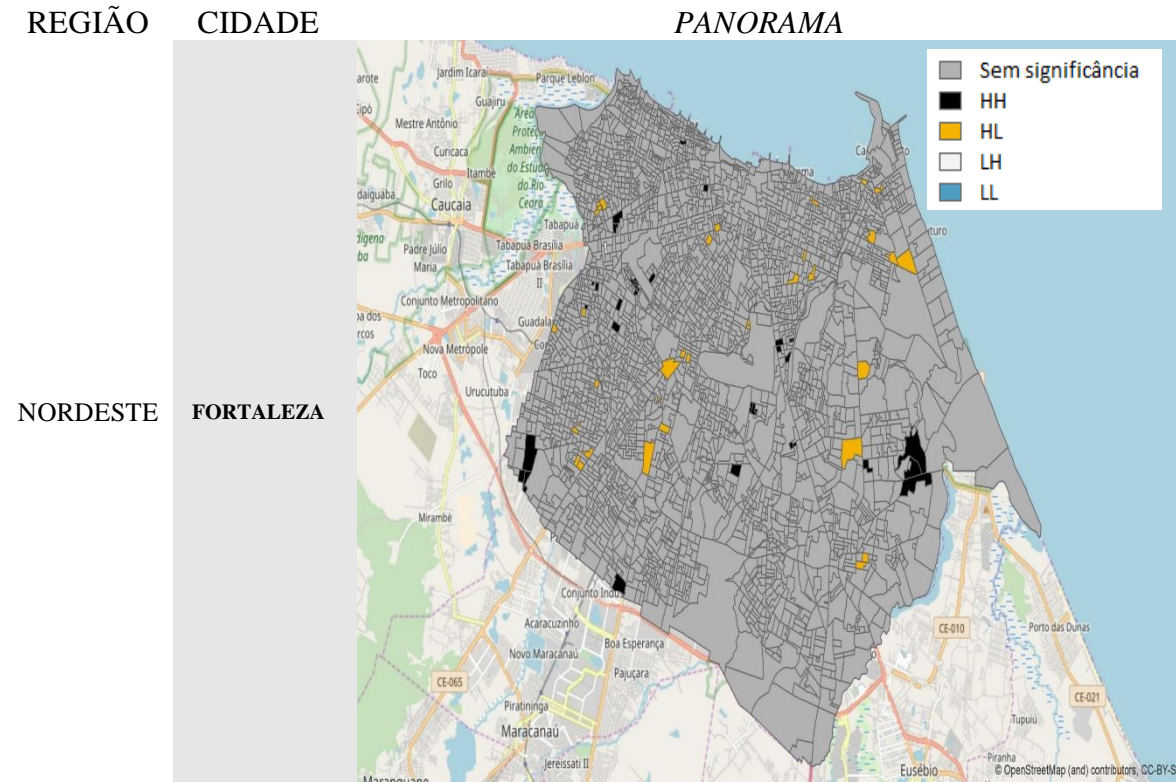


Fonte: Própria Autora.

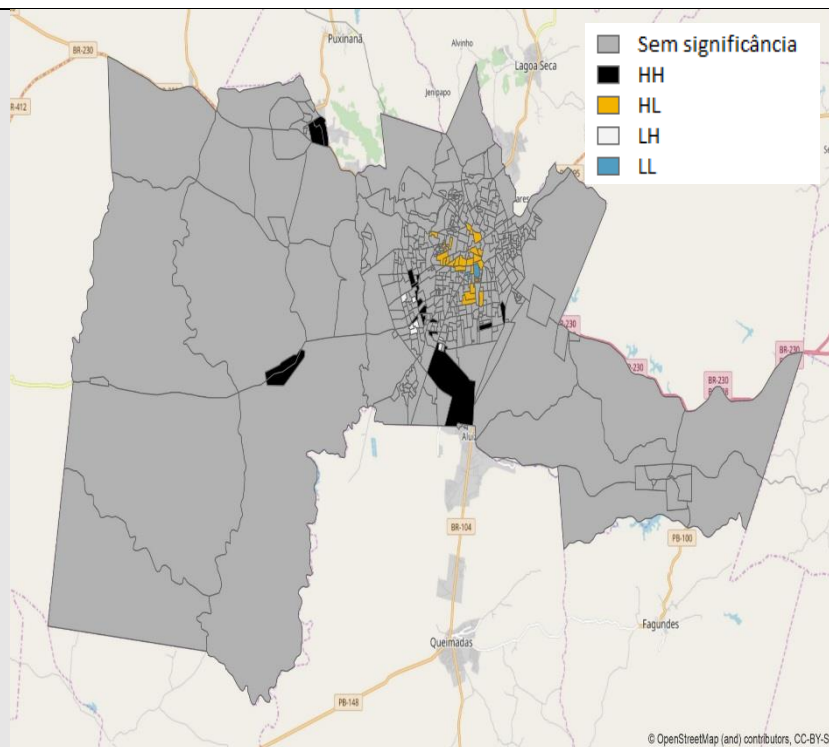
Apêndice G – Resultados da análise de *clusters* e *outliers*, de acordo com o índice de Anselin Local Moran, por cidade de cada região

O método de Anselin Local Moran é útil para compreendermos sobre o comportamento localizado das participações nos setores censitários. Os resultados de cada uma das cidades podem ser observados a seguir, no Quadro 8.

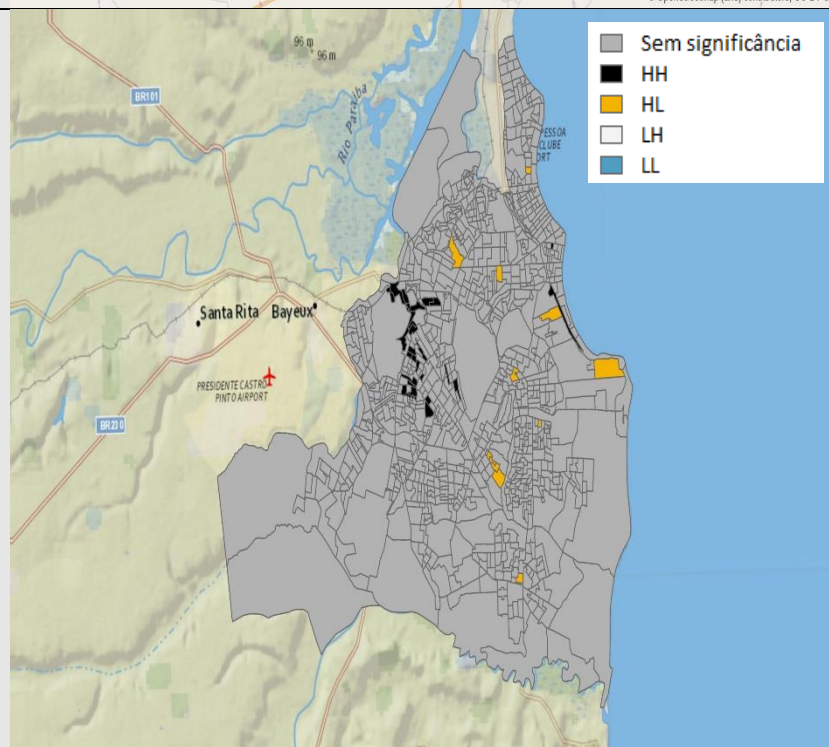
Quadro 8. Resultados da análise de *clusters* e *outliers*, sob enfoque do índice de Anselin Local Moran



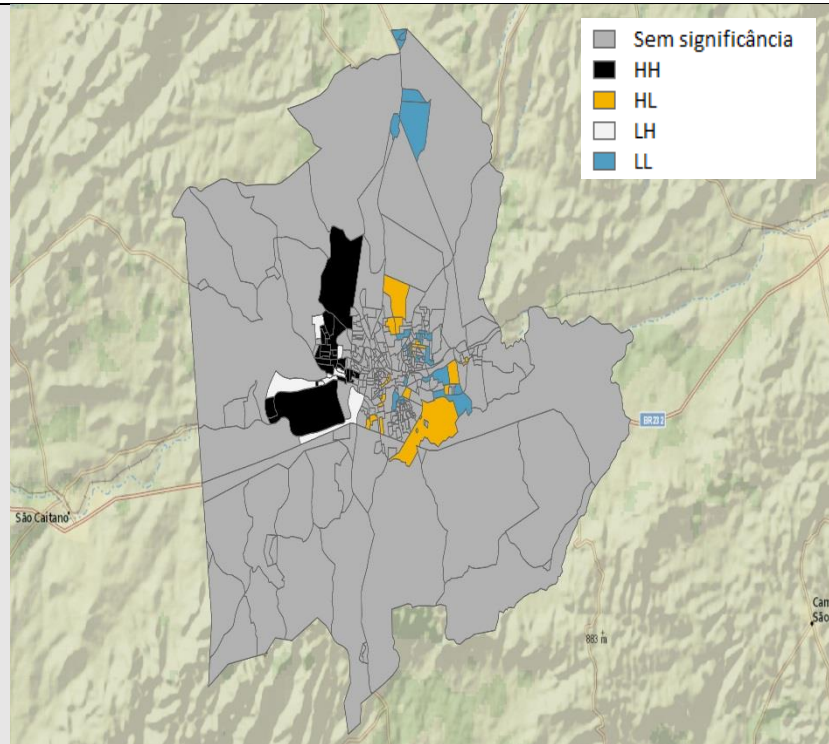
**CAMPINA
GRANDE**



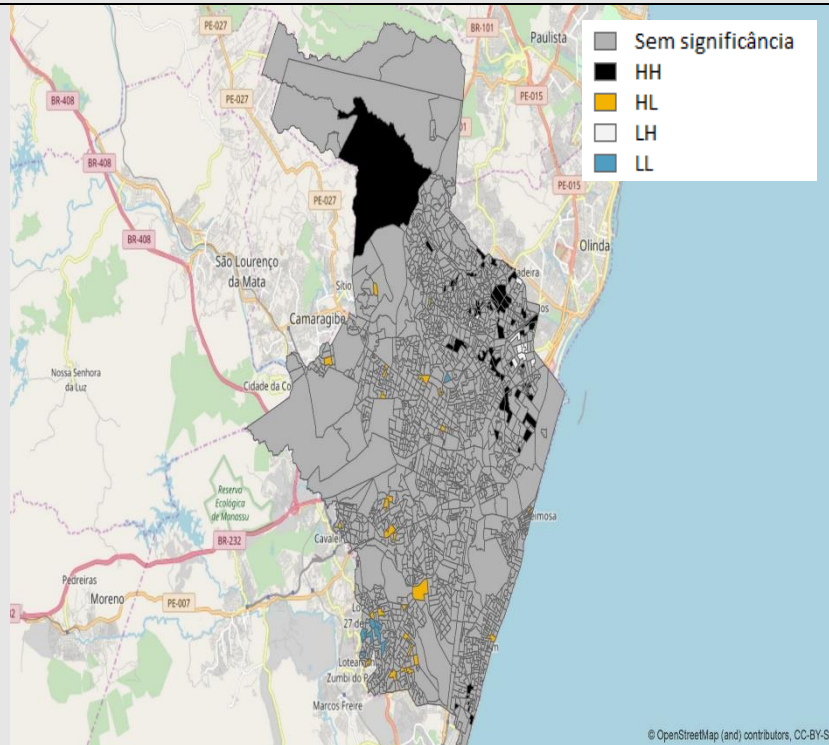
**JOÃO
PESSOA**



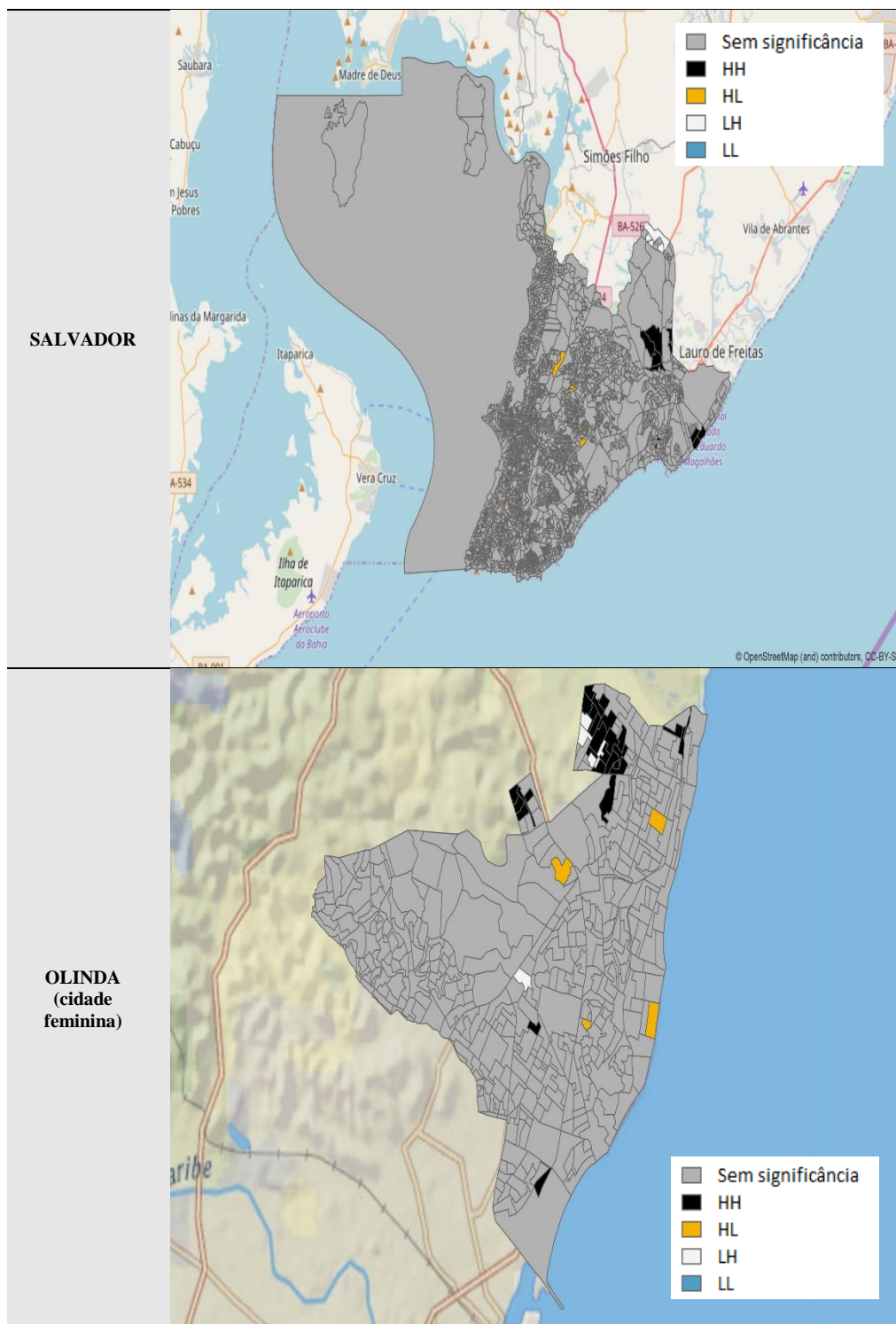
CARUARU

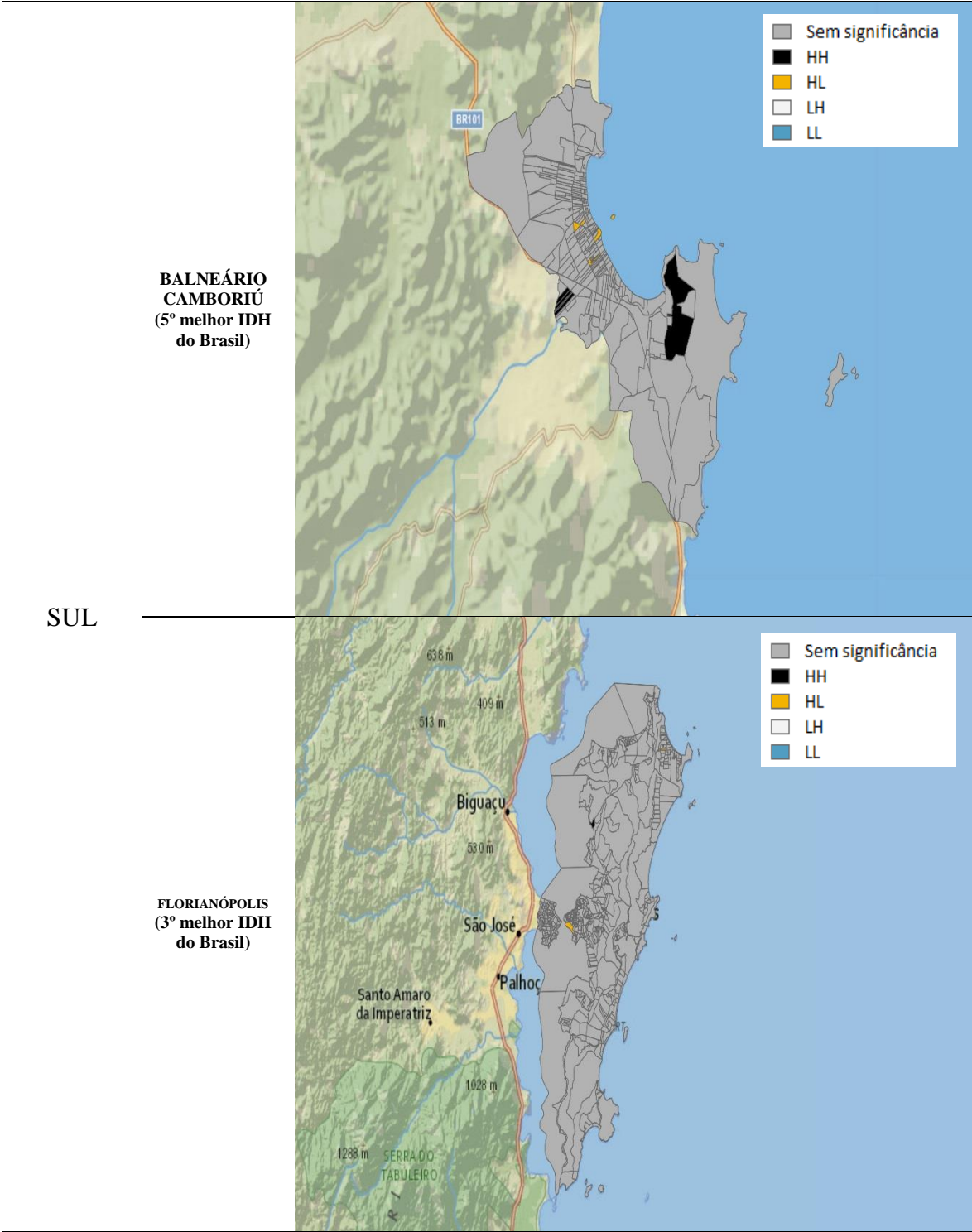


RECIFE (cidade feminina)

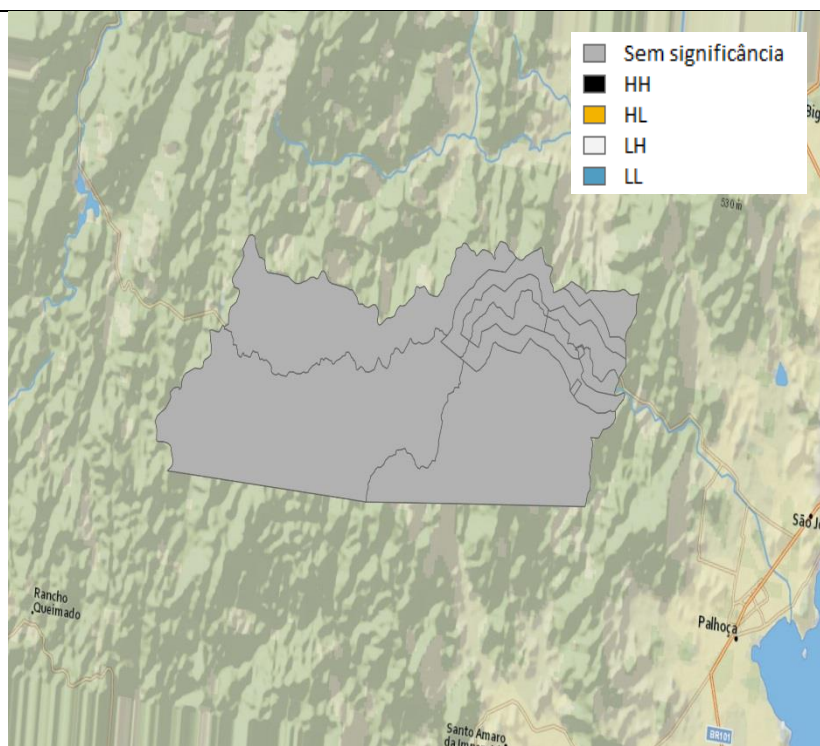


© OpenStreetMap (and) contributors, CC-BY-S



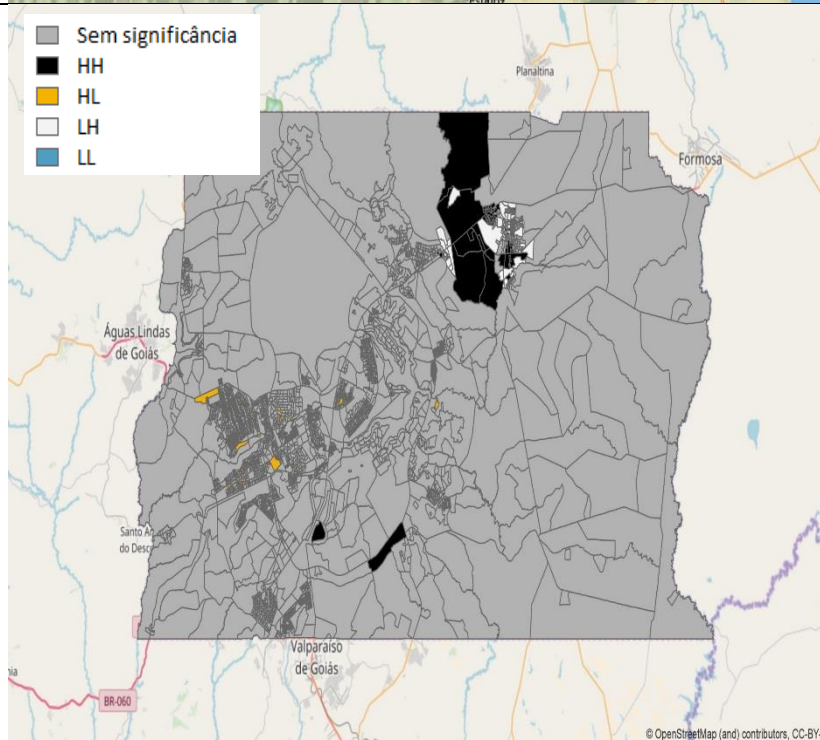


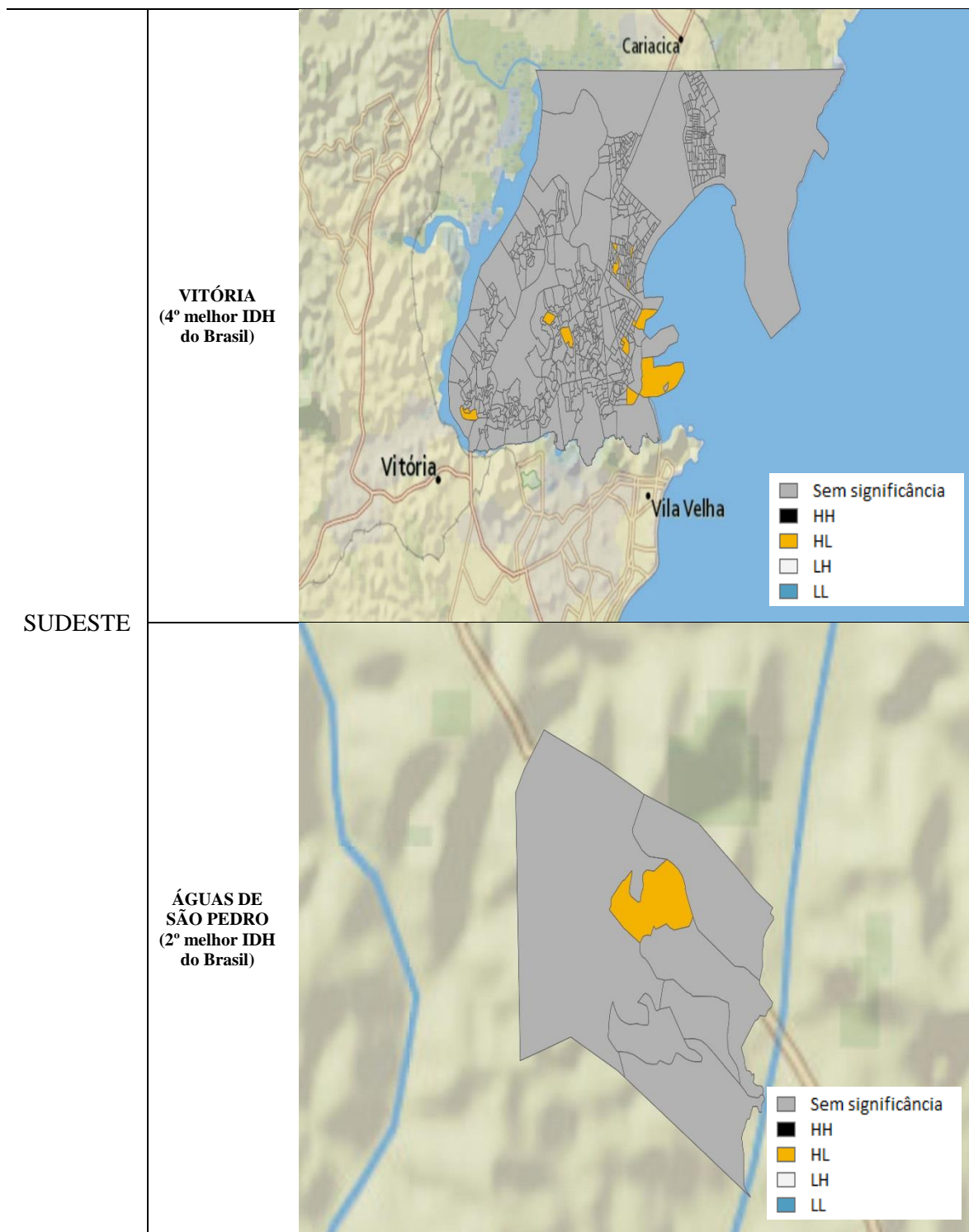
**SÃO PEDRO
DE
ALCÂNTARA
(cidade
masculina)**



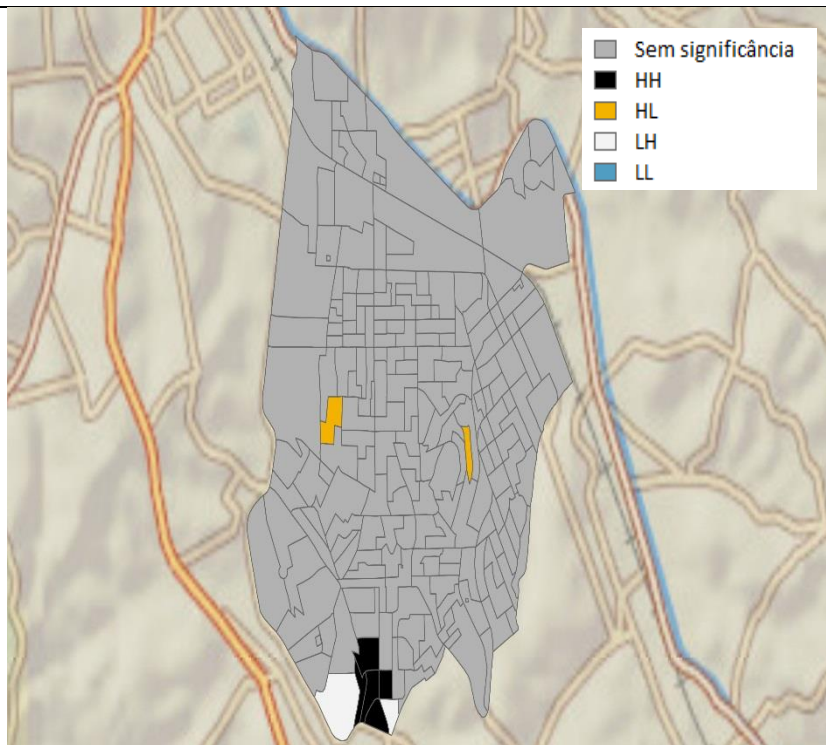
**CENTRO-
OESTE**

BRASÍLIA

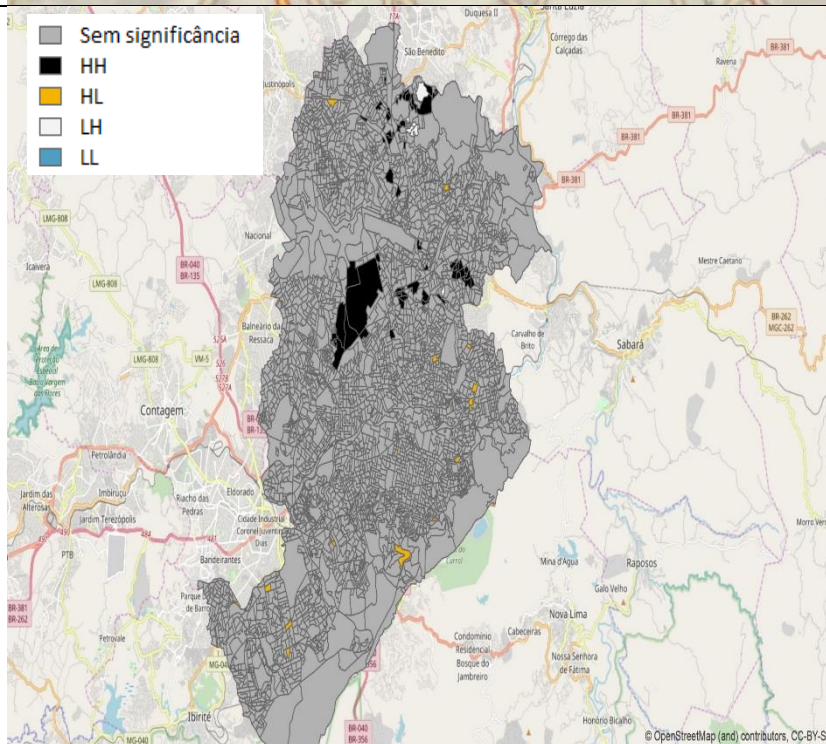




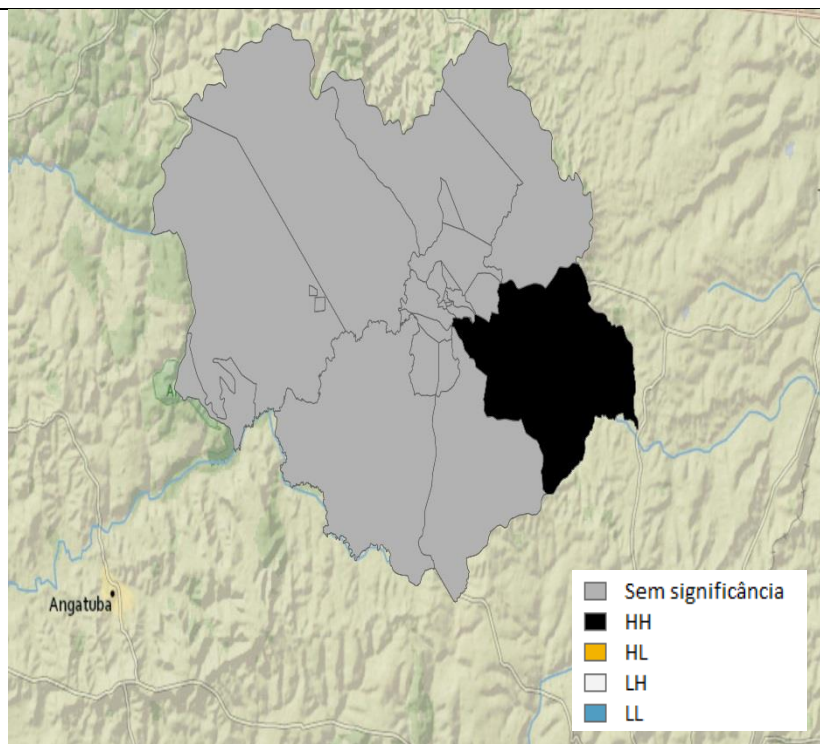
**SÃO
CAETANO
DO SUL**
(1º melhor IDH
do Brasil,
cidade
feminina)



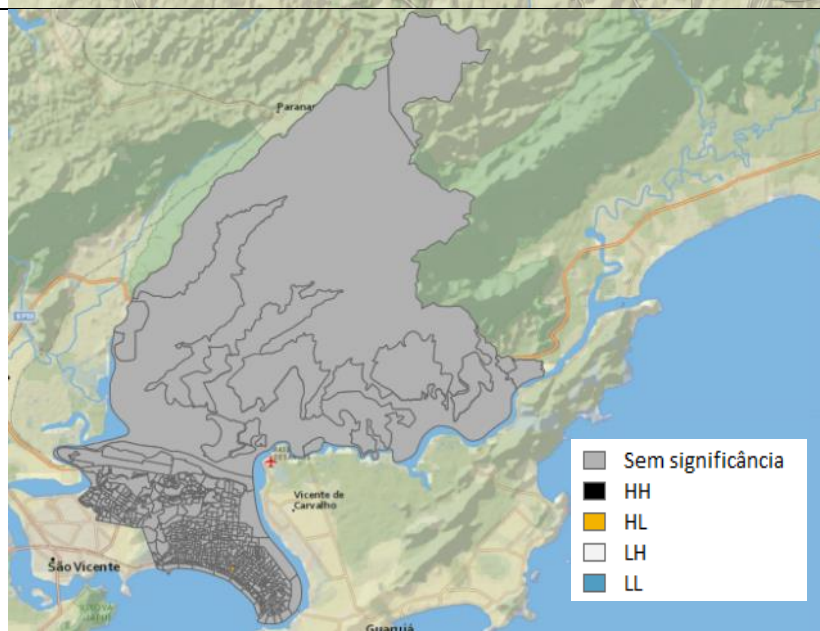
**BELO
HORIZONTE**



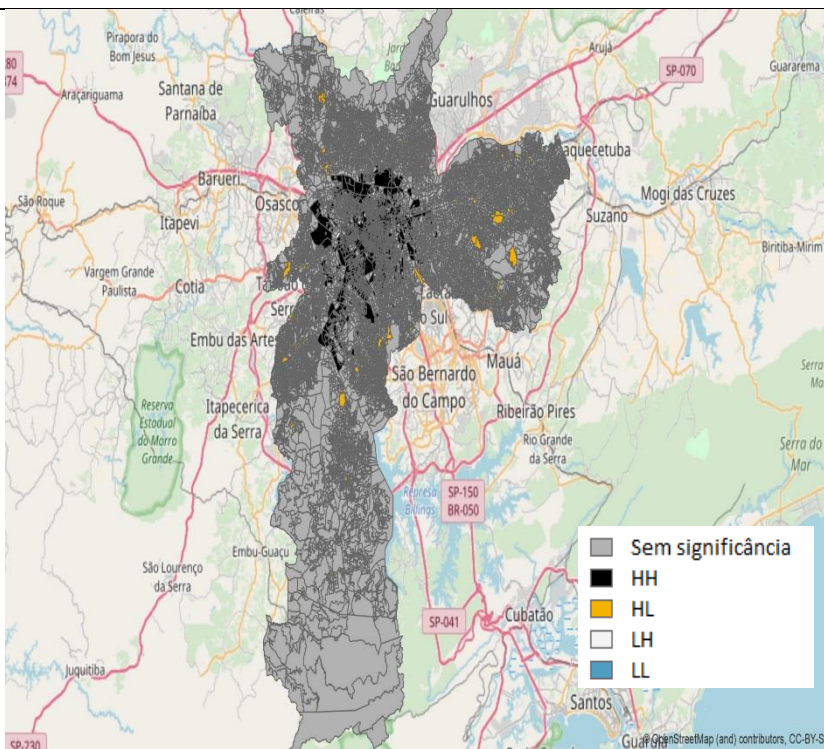
GUAREÍ
(cidade masculina)



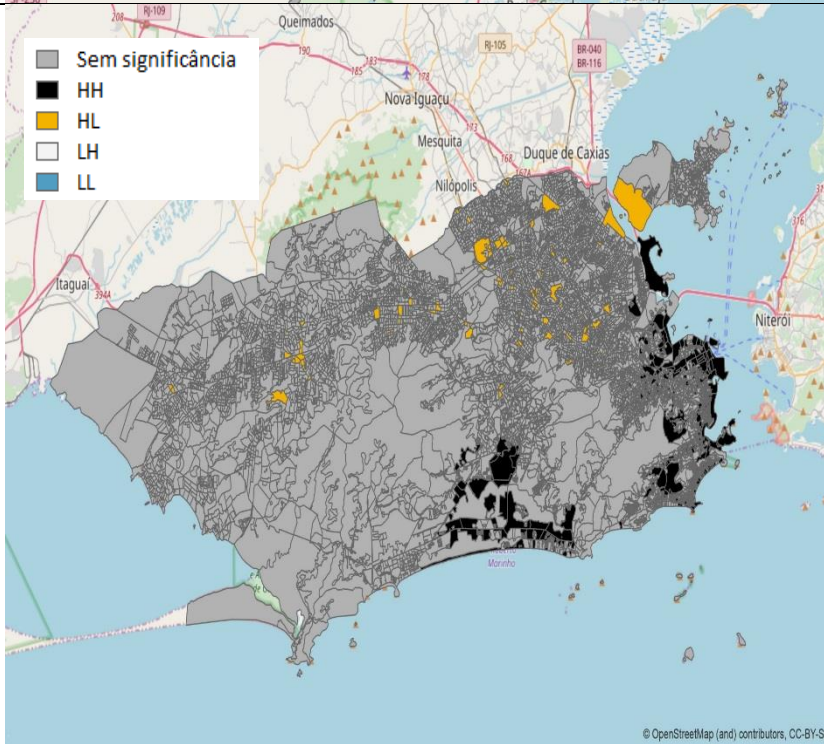
SANTOS
(cidade feminina)



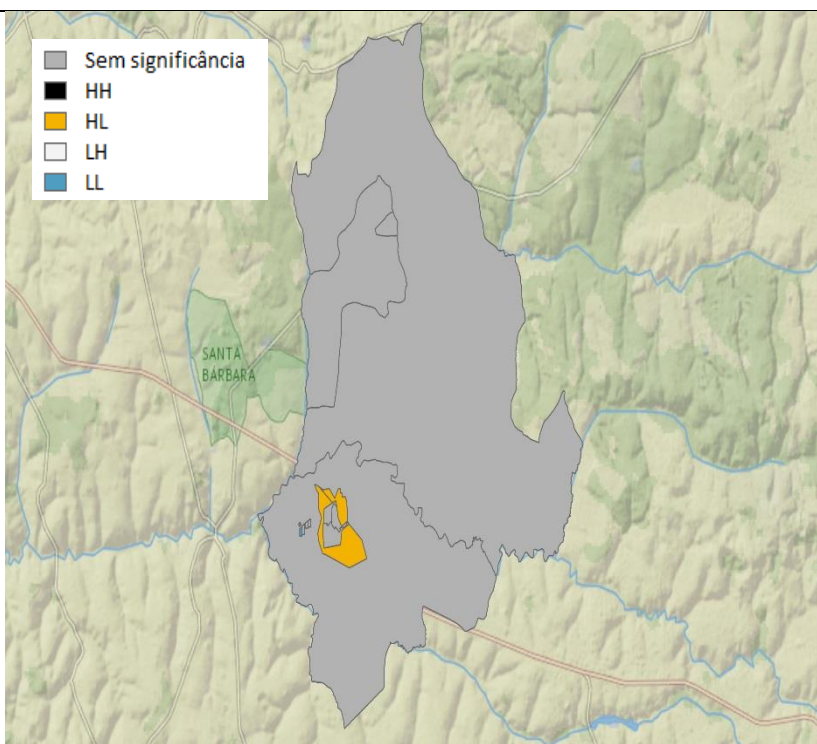
SÃO PAULO



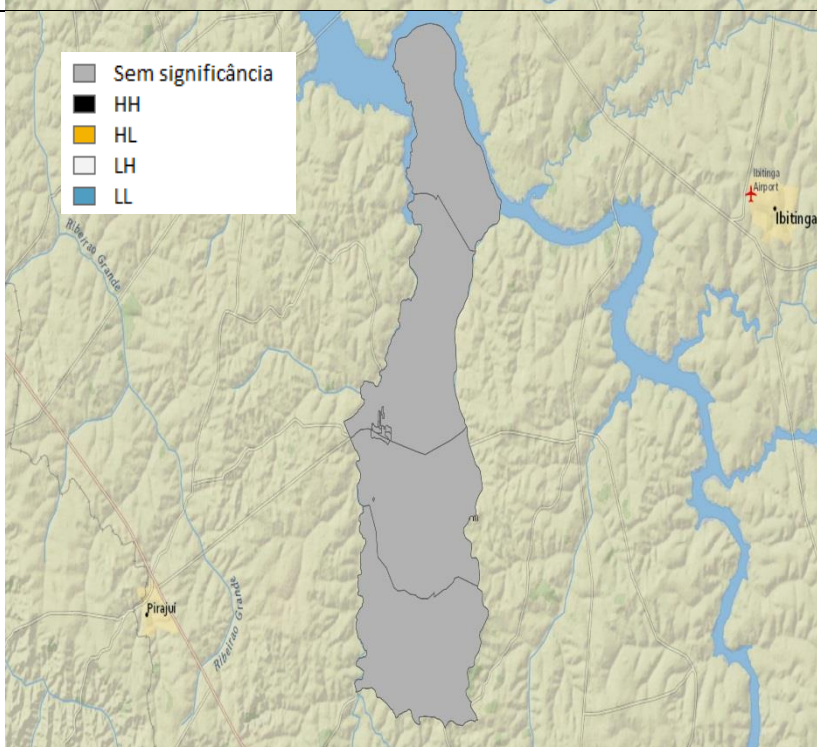
RIO DE JANEIRO

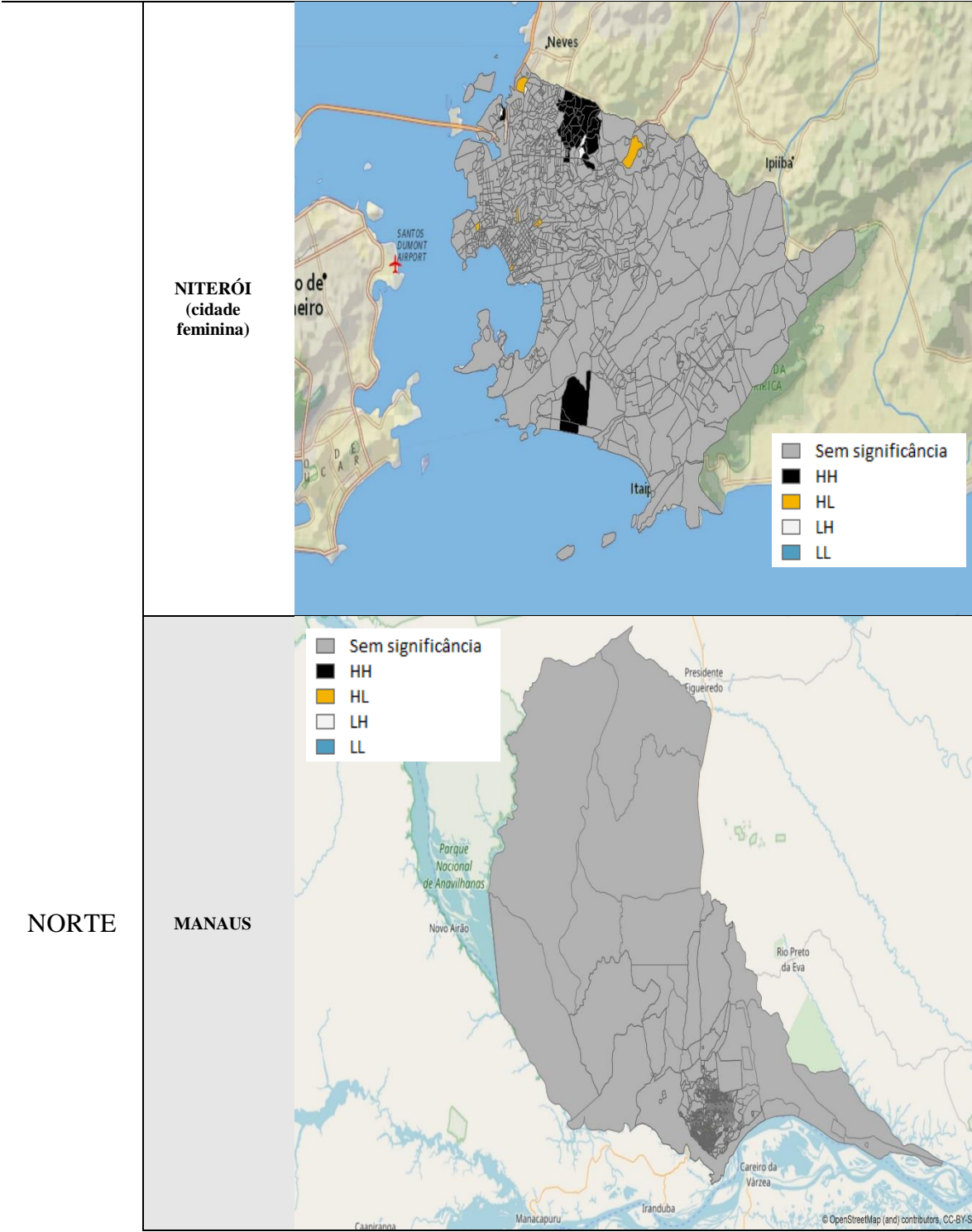


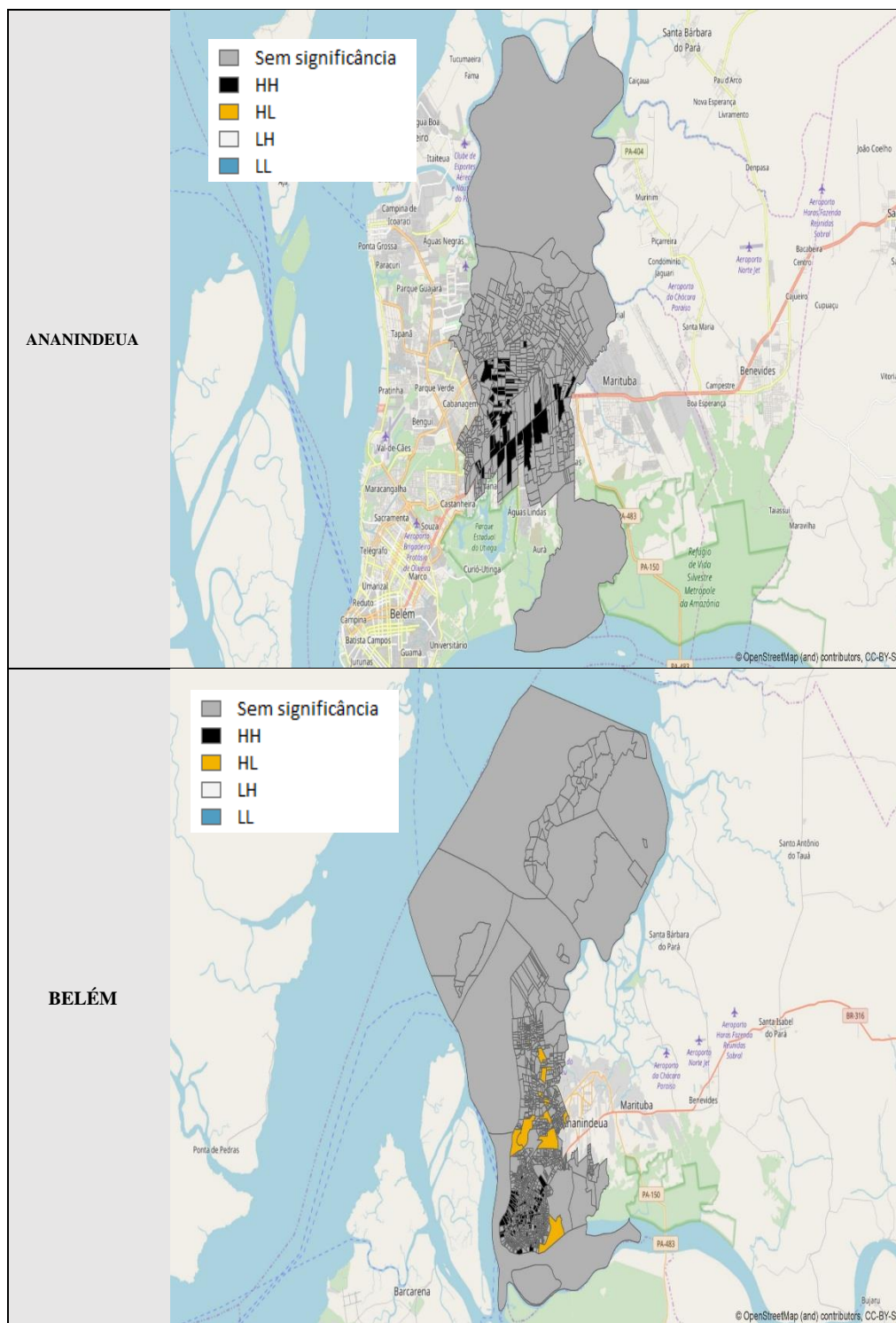
IARAS
(cidade masculina)



REGINÓPOLIS
(cidade masculina)





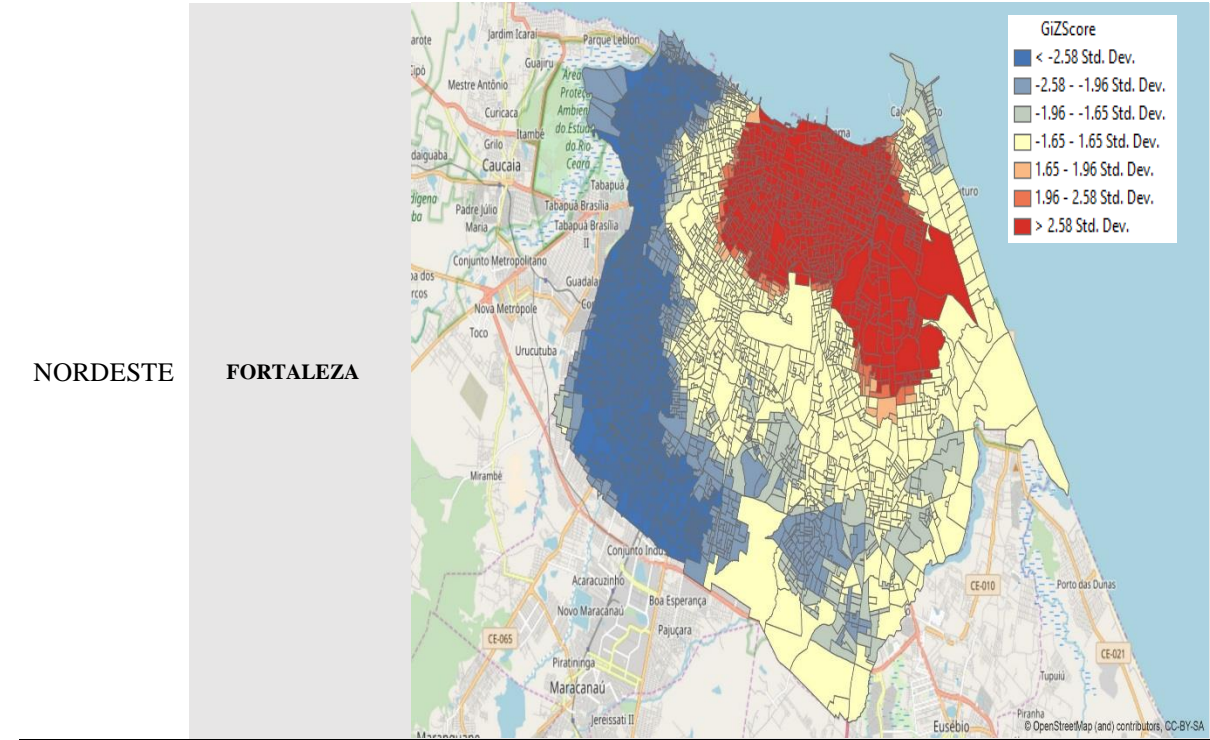


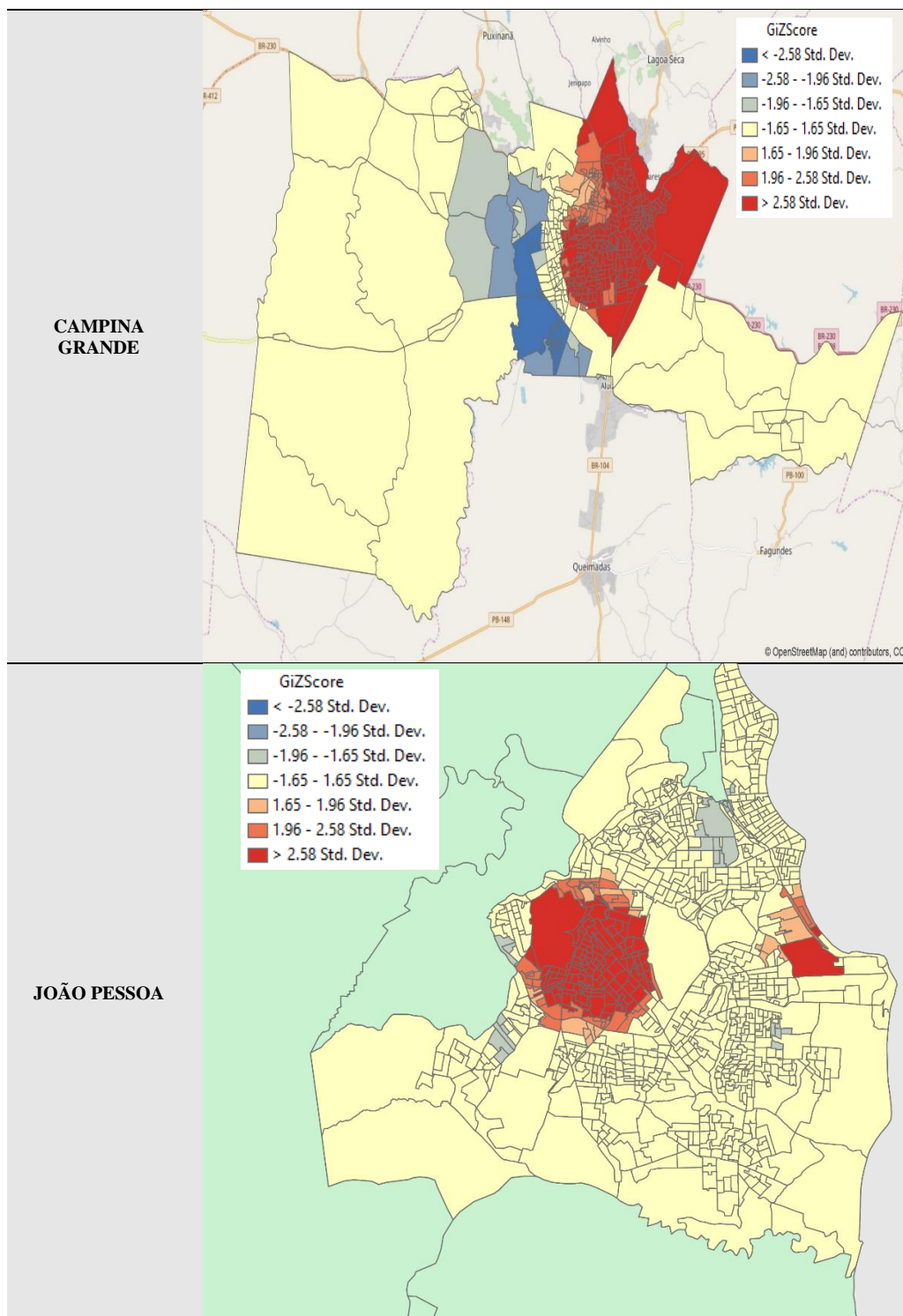
Fonte: Própria Autora.

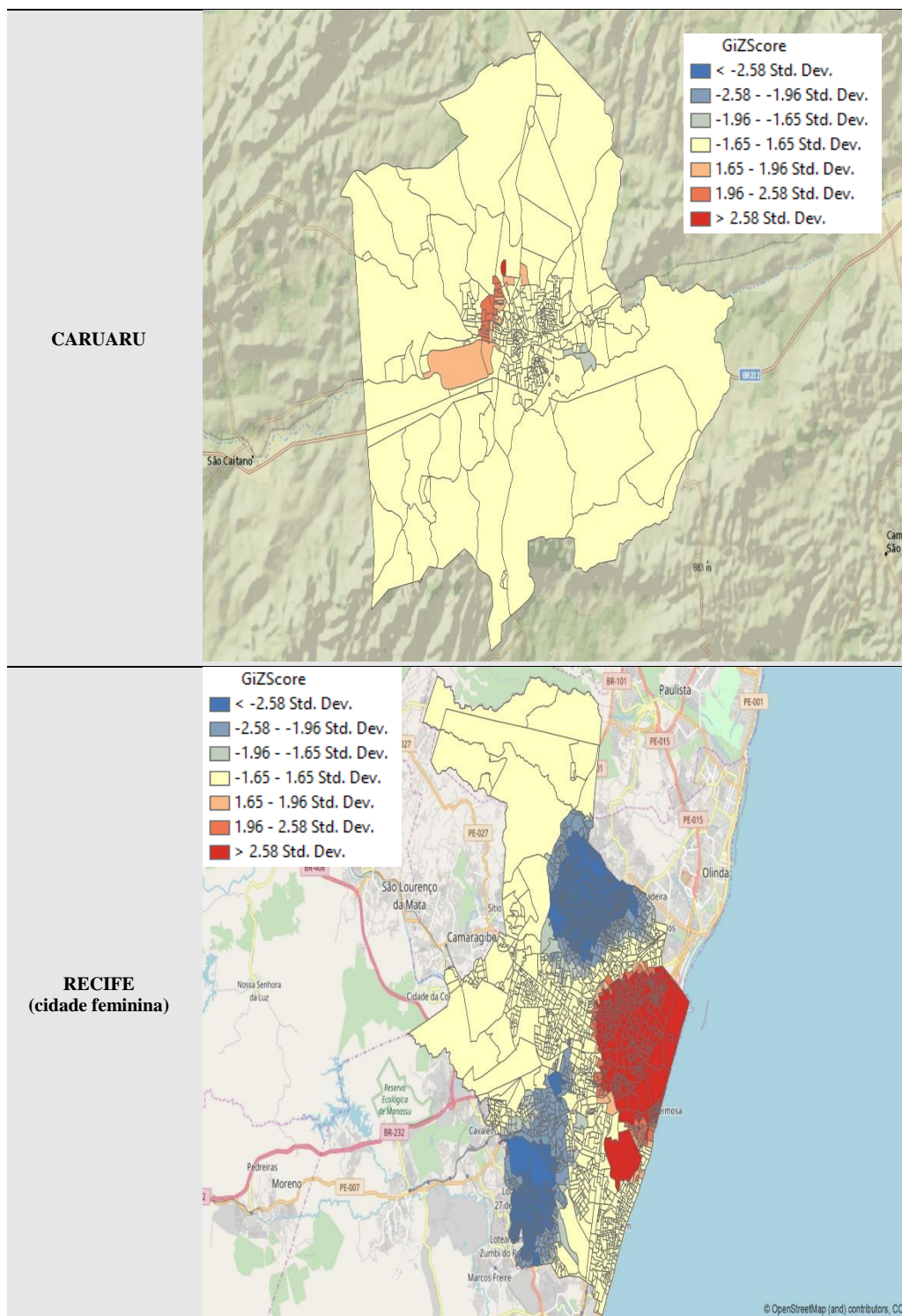
Apêndice H – Resultados da análise de áreas muito frequentadas (*Hotspots*), de acordo com a estatística de Getis-Ord G_i^*

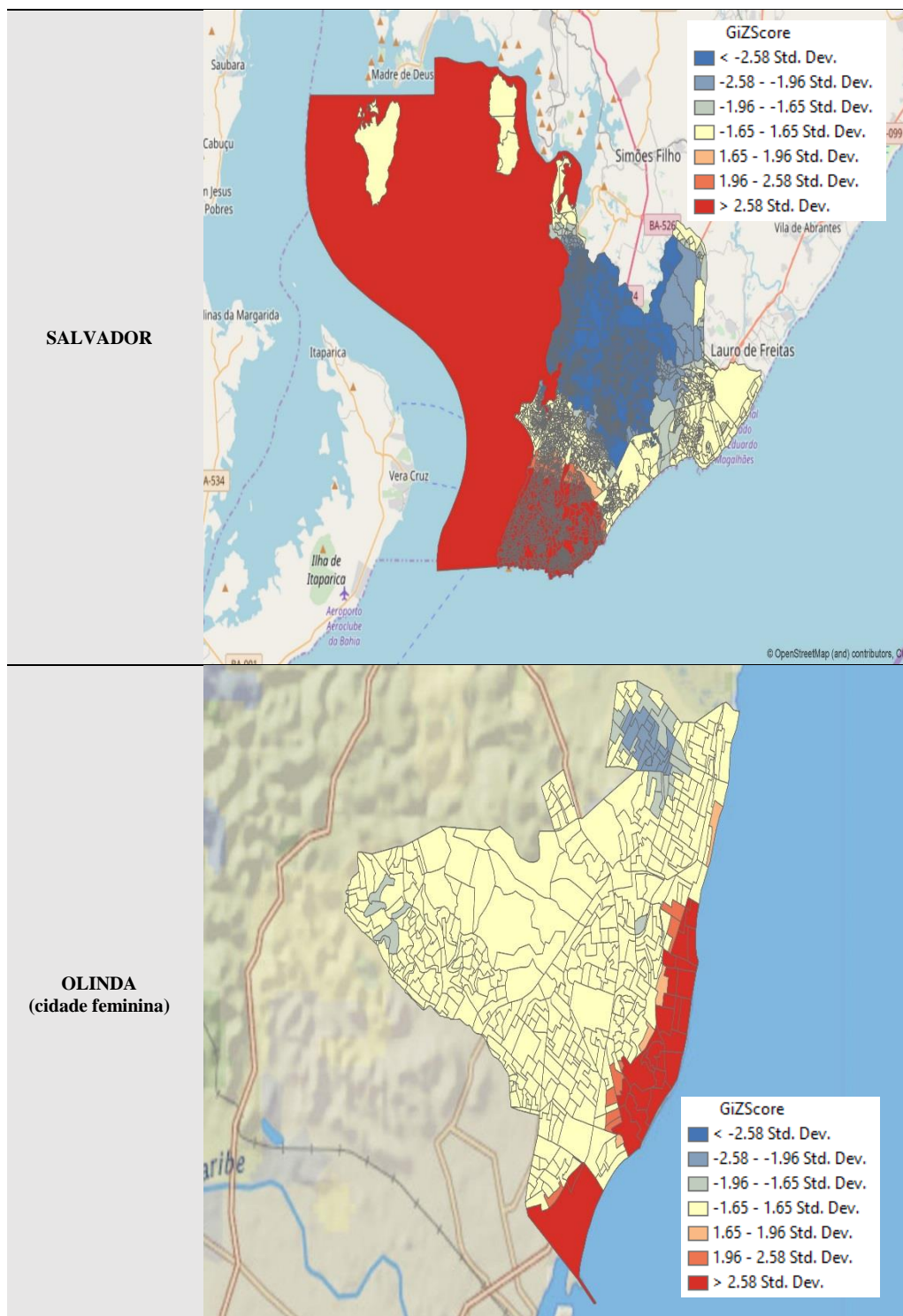
Neste momento, é observado quais áreas das cidades contém setores censitários que são muito ou pouco visitados. Os resultados podem ser observados no Quadro 9. No geral, as cidades selecionadas apresentaram esses dois tipos de áreas. Isso significa que muitas das cidades brasileiras possuem colaboradores que se dirigem para as mesmas áreas.

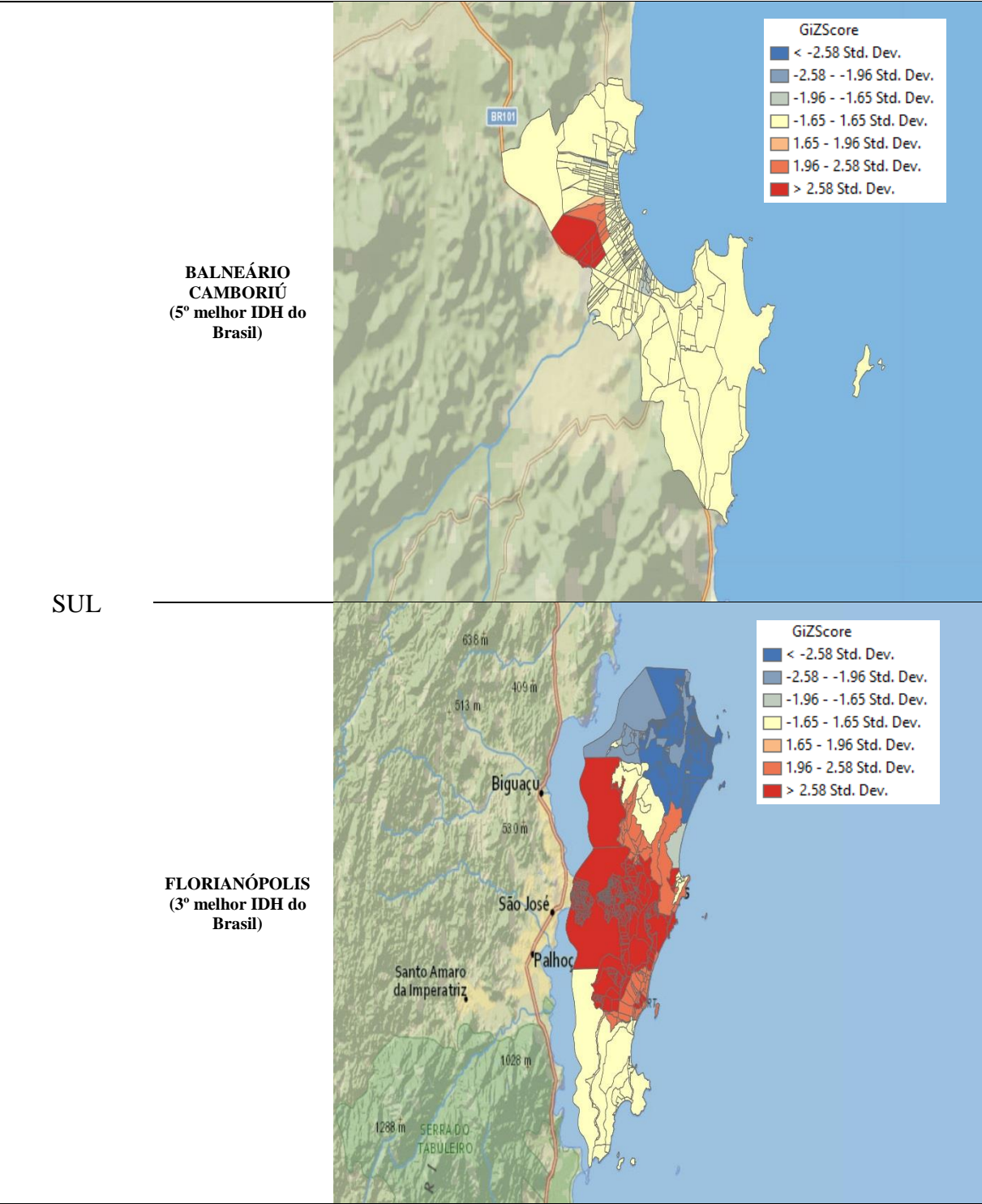
Quadro 9. Resultados da análise de *Hotspots* e *Coldspots* das cidades brasileiras selecionadas



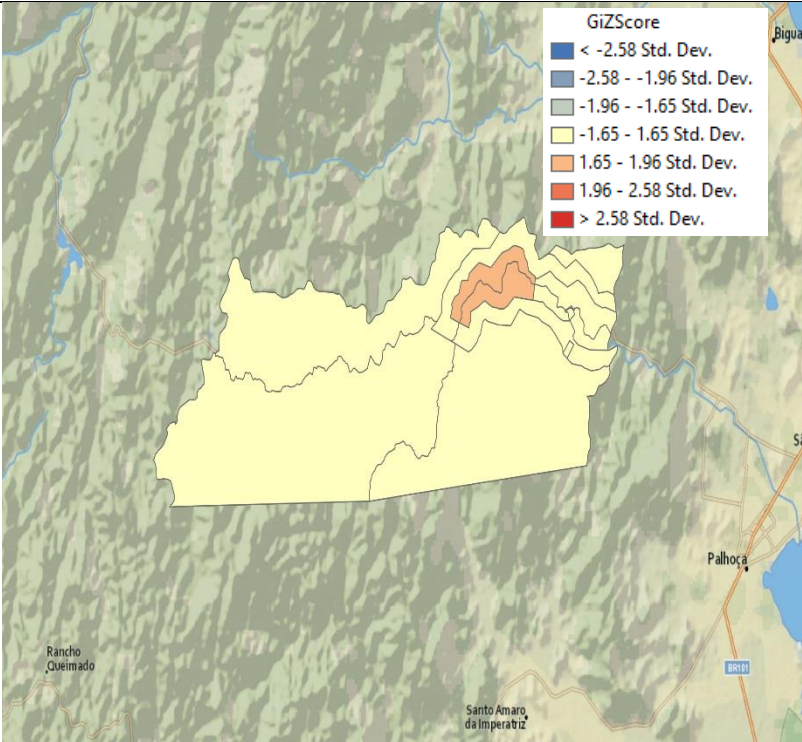






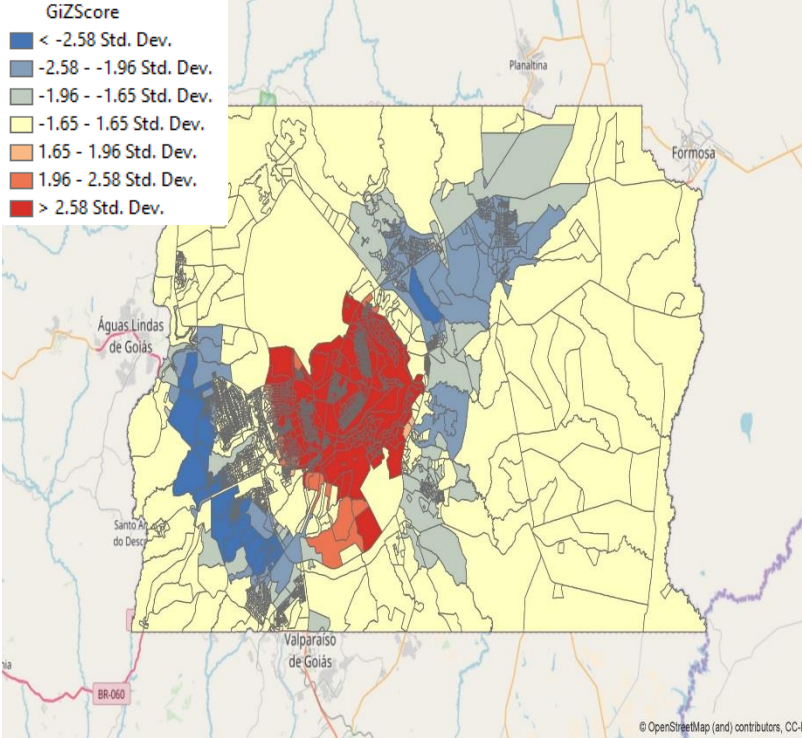


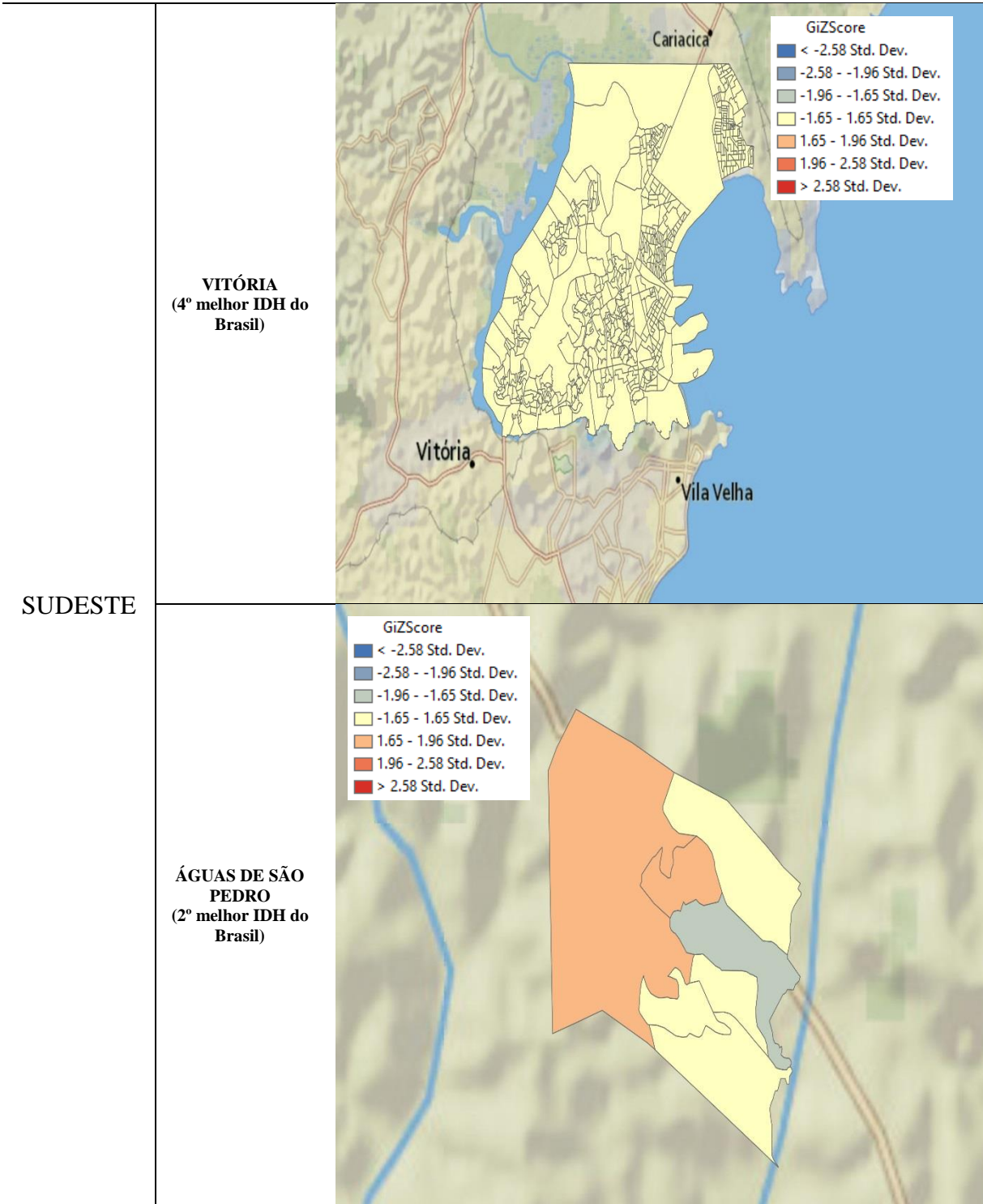
**SÃO PEDRO DE
ALCÂNTARA**
(cidade masculina)



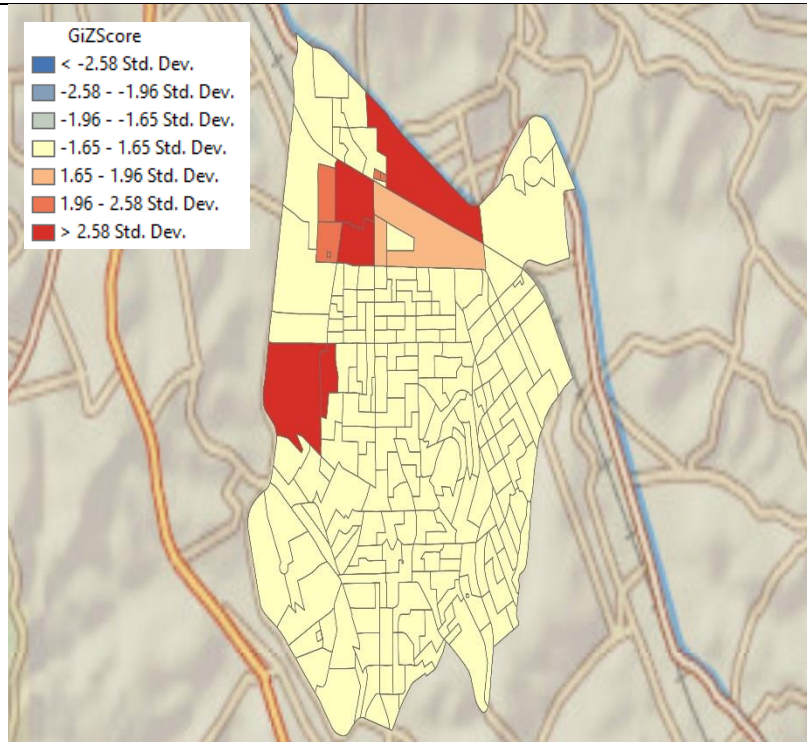
**CENTRO-
OESTE**

BRASÍLIA

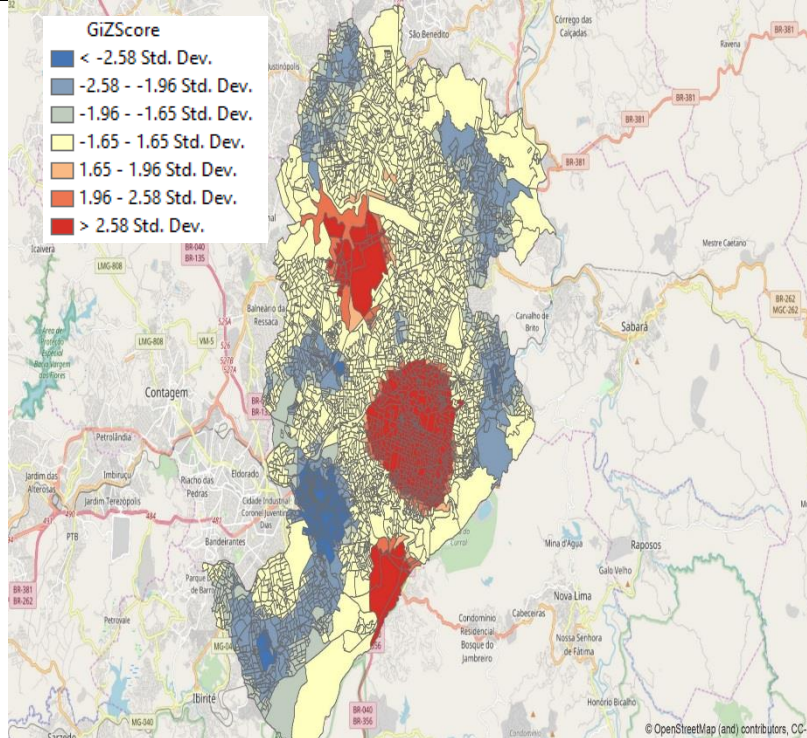




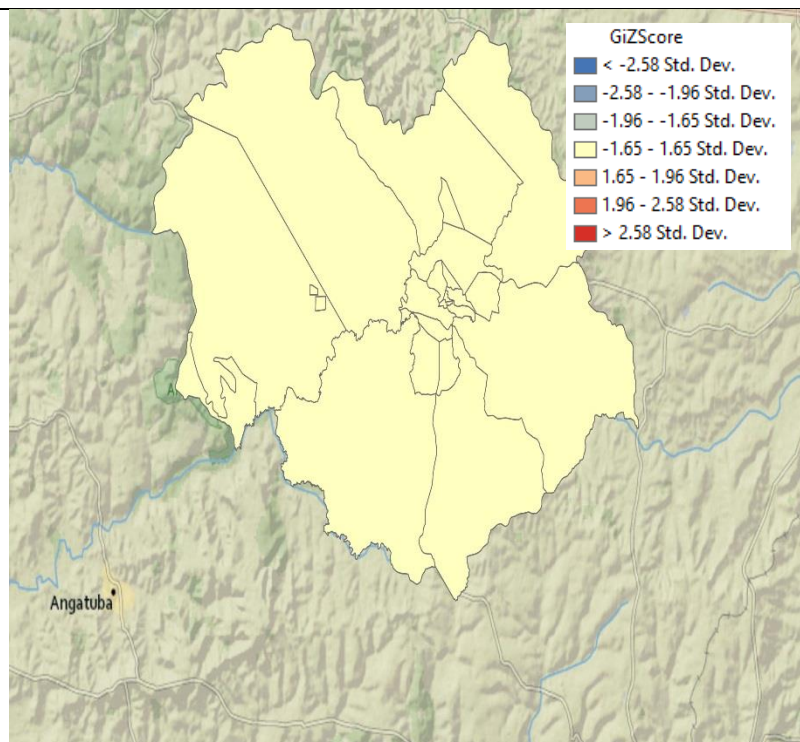
**SÃO CAETANO
DO SUL**
(1º melhor IDH do
Brasil, cidade
feminina)



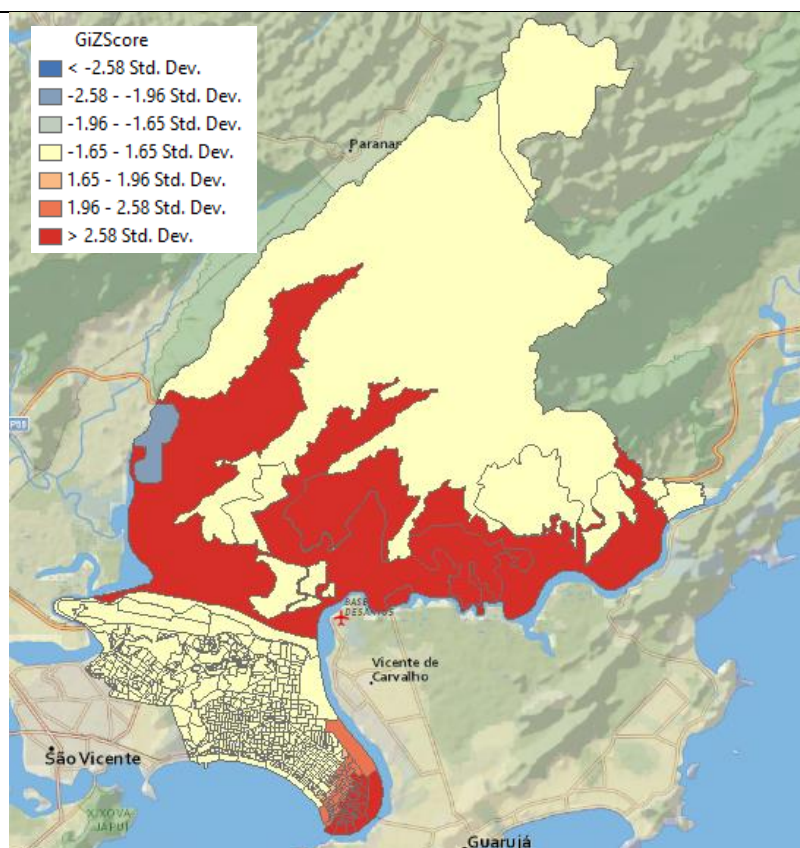
**BELO
HORIZONTE**



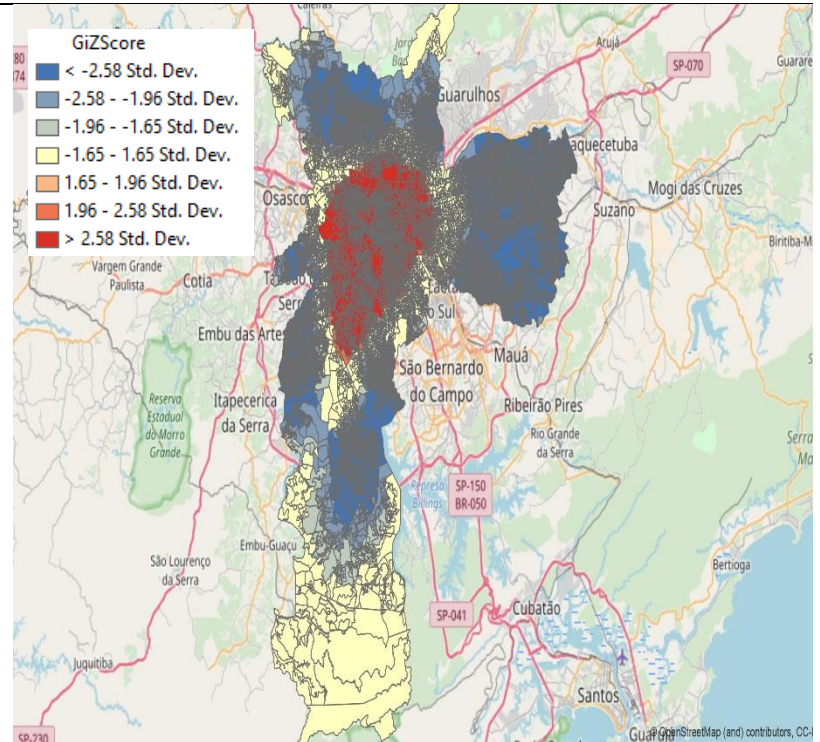
GUAREÍ
(cidade masculina)



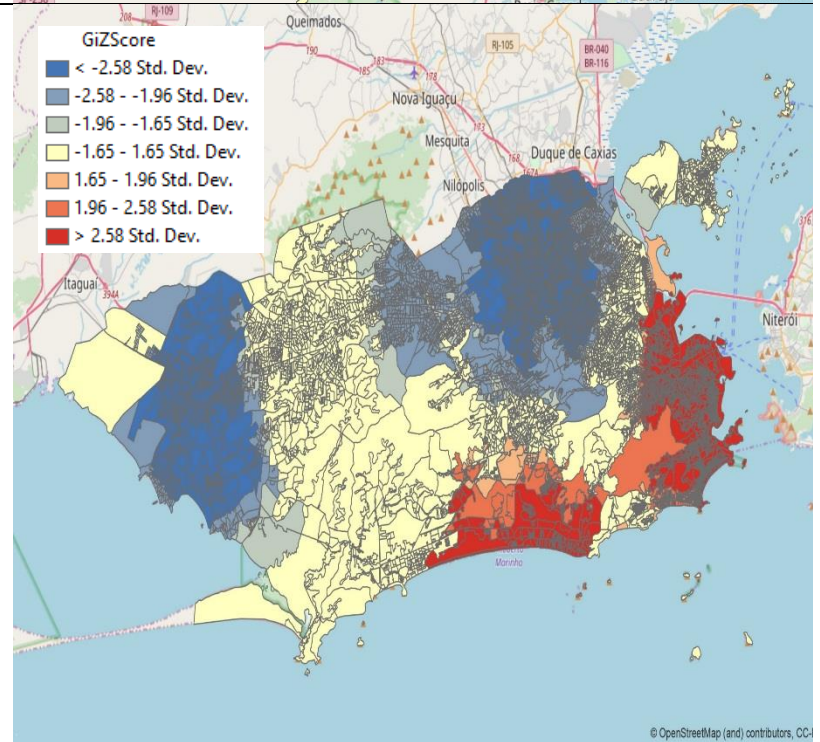
SANTOS
(cidade feminina)

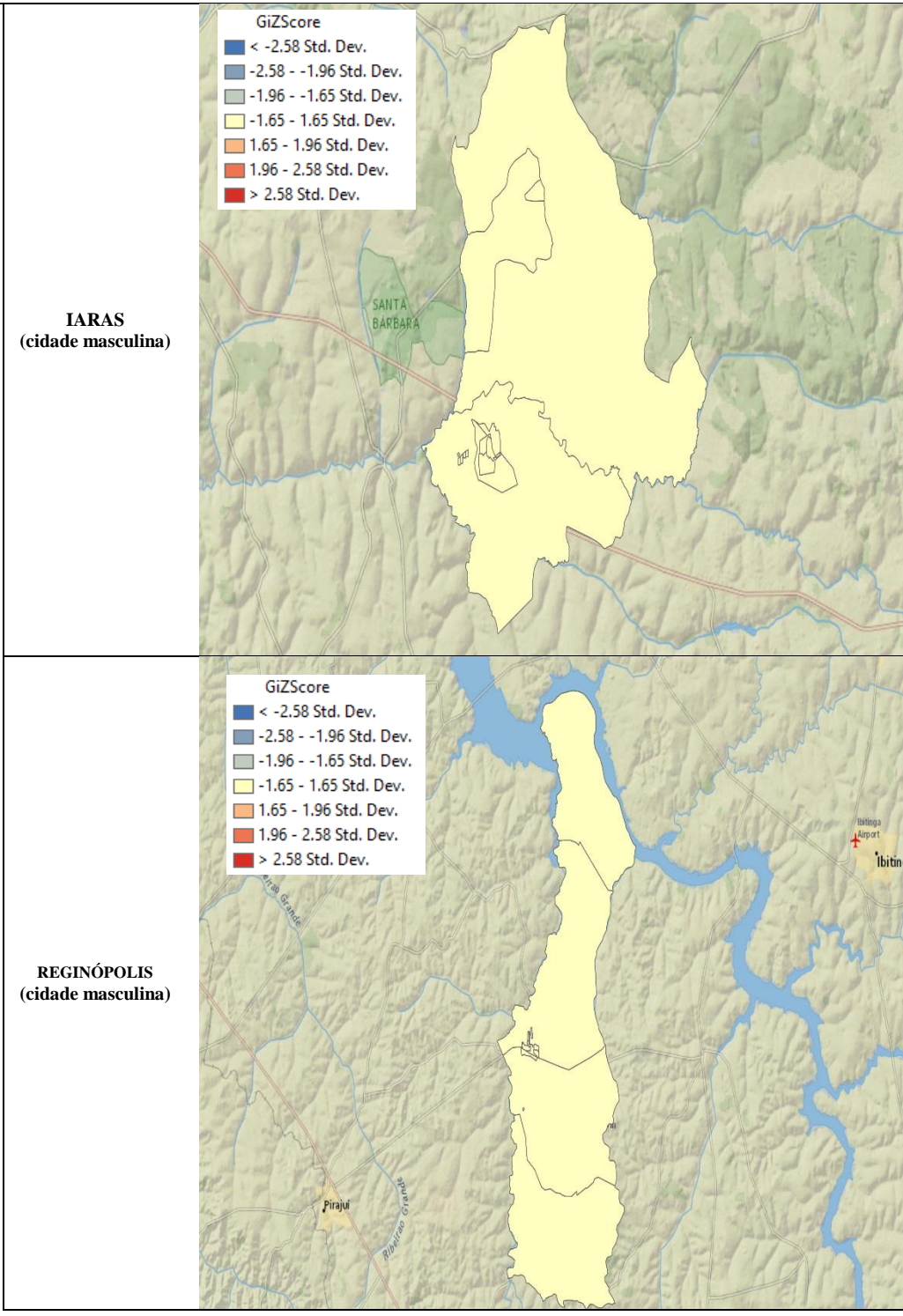


SÃO PAULO

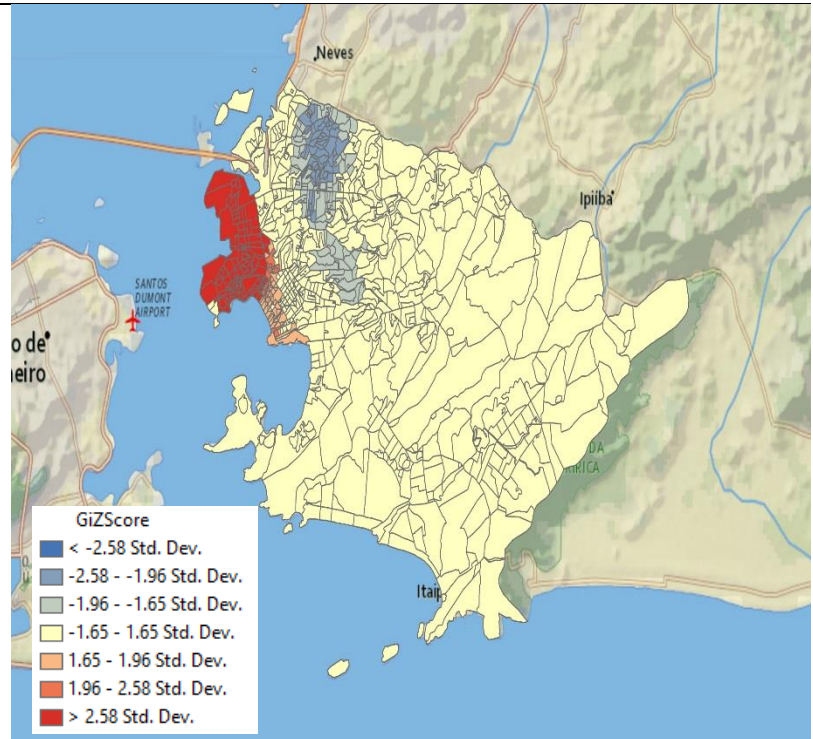


RIO DE JANEIRO



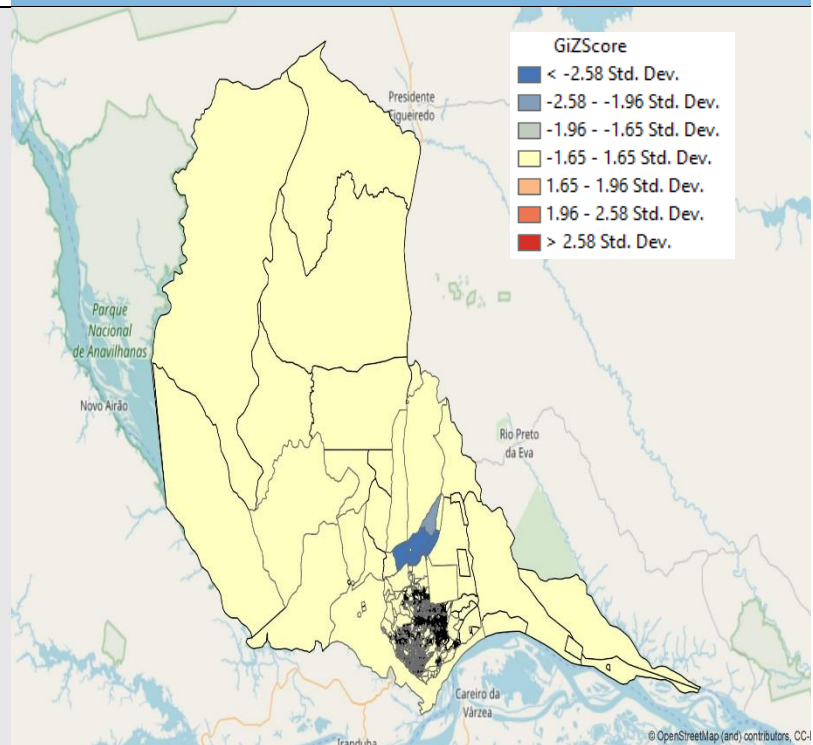


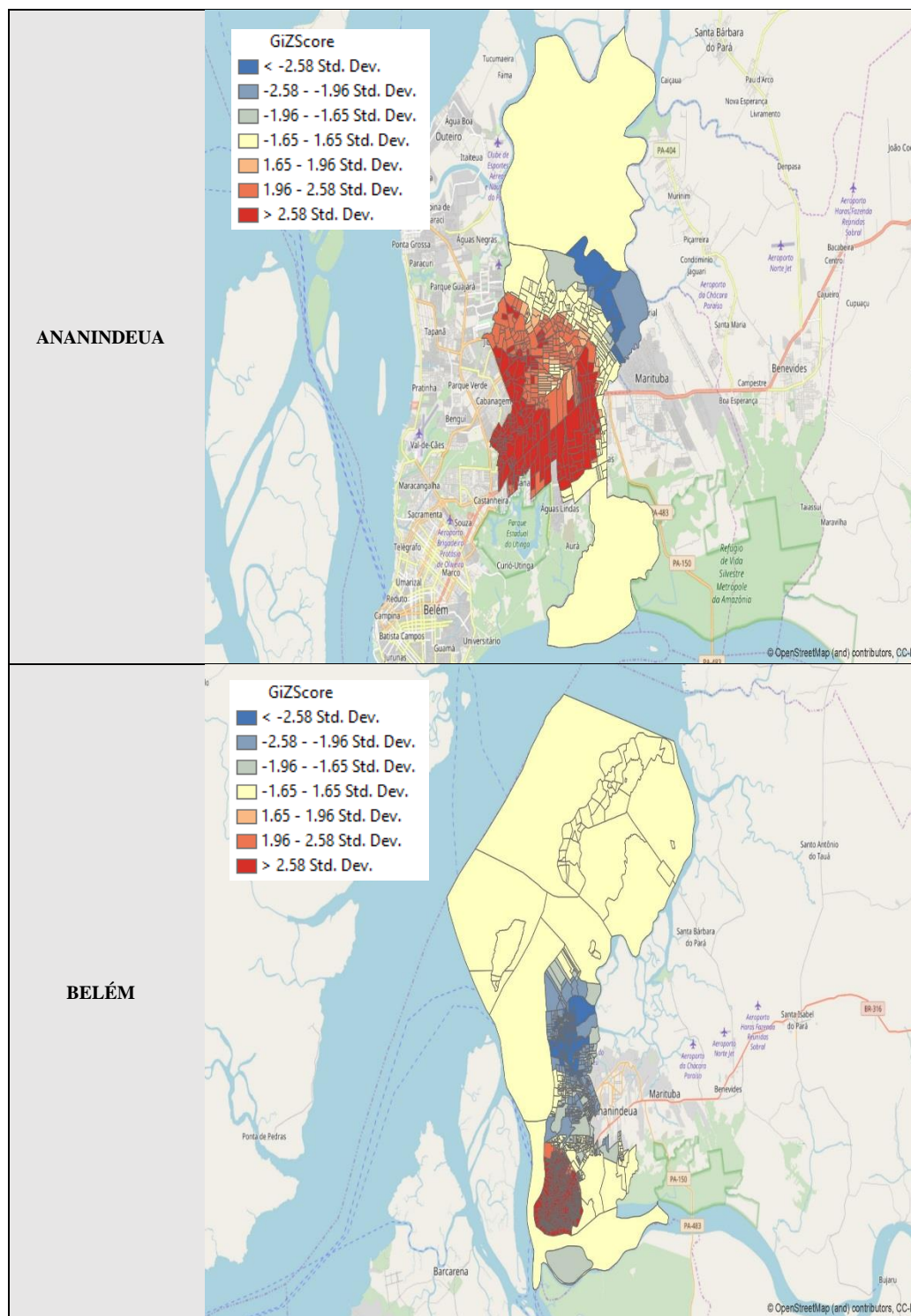
NITERÓI
(cidade feminina)



NORTE

MANAUS





Apêndice I – Fatores urbanos escolhidos para a análise de correlação

Existem muitas descrições sobre fatores urbanos. A presente tese selecionou descrições socioeconômicas sobre as unidades censitárias de cada cidade observada. No Quadro 10 estão descritas quais características urbanas definem cada temática selecionada.

Quadro 10. Características urbanas que descrevem cada setor censitário das cidades selecionadas

Temática da característica	Característica urbana
Tipo do setor censitário	<ul style="list-style-type: none">▪ Situação do setor censitário (urbana ou rural);
Tipos de domicílio	<ul style="list-style-type: none">▪ Número de domicílios particulares permanentes ou pessoas responsáveis por domicílios particulares permanentes;▪ Número de moradores em domicílios particulares permanentes ou população residente em domicílios particulares permanentes;▪ Média do número de moradores em domicílios particulares permanentes;▪ Variância do número de moradores em domicílios particulares permanentes;
Rendimento dos moradores do setor censitário	<ul style="list-style-type: none">▪ Valor do rendimento nominal médio mensal das pessoas responsáveis por domicílios particulares permanentes (com e sem rendimento);▪ Variância do rendimento nominal mensal das pessoas responsáveis por domicílios particulares permanentes (com e sem rendimento)▪ Valor do rendimento nominal médio mensal das pessoas responsáveis por domicílios particulares permanentes (com rendimento);▪ Variância do rendimento nominal mensal das pessoas responsáveis por domicílios particulares permanentes (com rendimento);▪ Valor do rendimento nominal médio mensal das pessoas de 10 anos ou mais de idade (com e sem rendimento);▪ Variância do rendimento nominal mensal das pessoas de 10 anos ou mais de idade (com e sem rendimento);▪ Valor do rendimento nominal médio mensal das pessoas de 10 anos ou mais de idade (com rendimento);▪ Variância do rendimento nominal mensal das pessoas de 10 anos

	ou mais de idade (com rendimento);
Infraestrutura do setor censitário	<ul style="list-style-type: none"> ▪ Número de domicílios particulares e domicílios coletivos; ▪ Número de domicílios particulares permanentes com lixo coletado; ▪ Número de domicílios particulares permanentes com energia elétrica; ▪ Número de domicílios particulares permanentes sem morador do sexo masculino; ▪ Número de domicílios particulares permanentes sem morador do sexo feminino; ▪ Número de domicílios particulares permanentes do tipo casa com abastecimento de água da rede geral; ▪ Número de domicílios particulares permanentes do tipo casa sem banheiro de uso exclusivo dos moradores; ▪ Número de domicílios particulares permanentes do tipo casa com lixo coletado; ▪ Número de domicílios particulares permanentes do tipo casa com energia elétrica; ▪ Número de domicílios particulares permanentes do tipo casa sem energia elétrica;
Características dos moradores do setor censitário	<ul style="list-style-type: none"> ▪ Número de pessoas alfabetizadas com 5 ou mais anos de idade; ▪ Número de pessoas alfabetizadas com 80 anos ou mais de idade; ▪ Número de mulheres alfabetizadas com 5 ou mais anos de idade; ▪ Número de pessoas residentes; ▪ Número de pessoas residentes e cor ou raça – branca; ▪ Número de pessoas residentes e cor ou raça – preta; ▪ Número de pessoas Residentes e cor ou raça – amarela; ▪ Número de pessoas Residentes e cor ou raça – parda; ▪ Número de pessoas Residentes e cor ou raça – indígena; ▪ Total de pessoas de até 10 anos de idade que tinham registro de nascimento; ▪ Total de pessoas de até 10 anos de idade que não tinham registro de nascimento; ▪ Total de pessoas de até 10 anos de idade que não sabiam se tinham registro de nascimento (inclusive sem declaração).

Fonte: Própria Autora.

Apêndice J – Fatores urbanos que afetam a intensidade de colaborações nos setores censitários das cidades brasileiras

Nesta seção, será descrito quais fatores urbanos influenciam a participação espacial dos criadores de conteúdo nos setores censitários, em cidades brasileiras. Mais detalhes no Quadro 11.

Quadro 11. Resumo da análise de correlação de Kendall entre a quantidade de colaborações e características urbanas, (+) indica correlação positiva e (-) correlação negativa

Cidade	Fatores que influenciam a participação nos setores censitários
ANANINDEUA (PA)	-
ÁGUAS DE SÃO PEDRO (SP)	<ul style="list-style-type: none"> Variância do número de moradores em domicílios particulares permanentes (+)
BALNEÁRIO CAMBORIU (SC)	-
BELÉM	-
BELO HORIZONTE (MG)	-
BRASILIA (DF)	<ul style="list-style-type: none"> Valor do rendimento nominal médio mensal das pessoas responsáveis por domicílios particulares permanentes (com e sem rendimento) (+) Valor do rendimento nominal médio mensal das pessoas responsáveis por domicílios particulares permanentes (com rendimento) (+) Valor do rendimento nominal médio mensal das pessoas de 10 anos ou mais de idade (com e sem rendimento) (+) Variância do rendimento nominal mensal das pessoas de 10 anos ou mais de idade (com e sem rendimento) (+) Valor do rendimento nominal médio mensal das pessoas de 10 anos ou mais de idade (com rendimento) (+)
CARUARU (PE)	<ul style="list-style-type: none"> Valor do rendimento nominal médio mensal das pessoas responsáveis por domicílios particulares permanentes (com e sem rendimento) (+) Variância do rendimento nominal mensal das pessoas responsáveis por domicílios particulares permanentes (com e sem rendimento) (+) Valor do rendimento nominal médio mensal das pessoas responsáveis por domicílios particulares permanentes (com rendimento) (+) Variância do rendimento nominal mensal das pessoas responsáveis por domicílios particulares permanentes (com rendimento) (+) Valor do rendimento nominal médio mensal das pessoas de 10 anos ou mais de idade (com e sem rendimento) (+) Variância do rendimento nominal mensal das pessoas de 10 anos ou mais de idade (com e sem rendimento) (+) Valor do rendimento nominal médio mensal das pessoas de 10 anos ou mais de idade (com rendimento) (+) Variância do rendimento nominal mensal das pessoas de 10 anos ou mais de idade (com rendimento) (+) Pessoas alfabetizadas com 80 anos ou mais de idade (+)
CAMPINA GRANDE (PB)	<ul style="list-style-type: none"> Valor do rendimento nominal médio mensal das pessoas responsáveis por domicílios particulares permanentes (com e sem rendimento) (+) Variância do rendimento nominal mensal das pessoas responsáveis por domicílios particulares permanentes (com e sem rendimento) (+)

	<ul style="list-style-type: none"> ▪ Valor do rendimento nominal médio mensal das pessoas responsáveis por domicílios particulares permanentes (com rendimento) (+) ▪ Variância do rendimento nominal mensal das pessoas responsáveis por domicílios particulares permanentes (com rendimento) (+) ▪ Valor do rendimento nominal médio mensal das pessoas de 10 anos ou mais de idade (com e sem rendimento) (+) ▪ Variância do rendimento nominal mensal das pessoas de 10 anos ou mais de idade (com e sem rendimento) (+) ▪ Valor do rendimento nominal médio mensal das pessoas de 10 anos ou mais de idade (com rendimento) (+) ▪ Variância do rendimento nominal mensal das pessoas de 10 anos ou mais de idade (com rendimento) (+)
FLORIANÓPOLIS (SC)	-
FORTALEZA (CE)	<ul style="list-style-type: none"> ▪ Valor do rendimento nominal médio mensal das pessoas responsáveis por domicílios particulares permanentes (com e sem rendimento) (+) ▪ Variância do rendimento nominal mensal das pessoas responsáveis por domicílios particulares permanentes (com e sem rendimento) (+) ▪ Valor do rendimento nominal médio mensal das pessoas responsáveis por domicílios particulares permanentes (com rendimento) (+) ▪ Variância do rendimento nominal mensal das pessoas responsáveis por domicílios particulares permanentes (com rendimento) (+) ▪ Valor do rendimento nominal médio mensal das pessoas de 10 anos ou mais de idade (com e sem rendimento) (+) ▪ Variância do rendimento nominal mensal das pessoas de 10 anos ou mais de idade (com e sem rendimento) (+) ▪ Valor do rendimento nominal médio mensal das pessoas de 10 anos ou mais de idade (com rendimento) (+) ▪ Variância do rendimento nominal mensal das pessoas de 10 anos ou mais de idade (com rendimento) (+) ▪ Pessoas alfabetizadas com 80 anos ou mais de idade (+)
GUAREÍ (SP)	<ul style="list-style-type: none"> ▪ Domicílios particulares permanentes ou pessoas responsáveis por domicílios particulares permanentes (+) ▪ Moradores em domicílios particulares permanentes ou população residente em domicílios particulares permanentes (+) ▪ Valor do rendimento nominal médio mensal das pessoas de 10 anos ou mais de idade (com e sem rendimento) (+) ▪ Domicílios particulares permanentes com lixo coletado (+) ▪ Domicílios particulares permanentes sem morador do sexo masculino (+) ▪ Domicílios particulares permanentes do tipo casa com lixo coletado (+) ▪ Domicílios particulares permanentes do tipo casa com energia elétrica (+) ▪ Mulheres alfabetizadas com 5 ou mais anos de idade (+)
IARAS (SP)	<ul style="list-style-type: none"> ▪ Domicílios particulares permanentes do tipo casa sem energia elétrica (+)
JOÃO PESSOA (PB)	<ul style="list-style-type: none"> ▪ Moradores em domicílios particulares permanentes ou população residente em domicílios particulares permanentes (+) ▪ Valor do rendimento nominal médio mensal das pessoas responsáveis por domicílios particulares permanentes (com e sem rendimento) (+) ▪ Variância do rendimento nominal mensal das pessoas responsáveis por domicílios particulares permanentes (com e sem rendimento) (+)

	<ul style="list-style-type: none"> ▪ Valor do rendimento nominal médio mensal das pessoas responsáveis por domicílios particulares permanentes (com rendimento) (+) ▪ Valor do rendimento nominal médio mensal das pessoas de 10 anos ou mais de idade (com e sem rendimento) (+) ▪ Variância do rendimento nominal mensal das pessoas de 10 anos ou mais de idade (com e sem rendimento) (+) ▪ Valor do rendimento nominal médio mensal das pessoas de 10 anos ou mais de idade (com rendimento) (+)
MANAUS	-
NITERÓI (RJ)	-
OLINDA (PE)	-
RECIFE (PE)	<ul style="list-style-type: none"> ▪ Valor do rendimento nominal médio mensal das pessoas responsáveis por domicílios particulares permanentes (com e sem rendimento) (+) ▪ Variância do rendimento nominal mensal das pessoas responsáveis por domicílios particulares permanentes (com e sem rendimento) (+) ▪ Valor do rendimento nominal médio mensal das pessoas responsáveis por domicílios particulares permanentes (com rendimento) (+) ▪ Variância do rendimento nominal mensal das pessoas responsáveis por domicílios particulares permanentes (com rendimento) (+) ▪ Valor do rendimento nominal médio mensal das pessoas de 10 anos ou mais de idade (com e sem rendimento) (+) ▪ Variância do rendimento nominal mensal das pessoas de 10 anos ou mais de idade (com e sem rendimento) (+) ▪ Valor do rendimento nominal médio mensal das pessoas de 10 anos ou mais de idade (com rendimento) (+) ▪ Variância do rendimento nominal mensal das pessoas de 10 anos ou mais de idade (com rendimento) (+) ▪ Pessoas alfabetizadas com 80 anos ou mais de idade (+)
REGINÓPOLIS (SP)	<ul style="list-style-type: none"> ▪ Domicílios particulares permanentes sem morador do sexo masculino (+) ▪ Pessoas alfabetizadas com 80 anos ou mais de idade (+)
RIO DE JANEIRO (RJ)	-
SALVADOR (SA)	-
SANTOS (SP)	-
SÃO PEDRO DE ALCÂNTARA (SP)	<ul style="list-style-type: none"> ▪ Situação do setor: urbana [-], rural [+] (-) ▪ Variância do número de moradores em domicílios particulares permanentes (-) ▪ Variância do rendimento nominal mensal das pessoas responsáveis por domicílios particulares permanentes (com e sem rendimento) (-) ▪ Variância do rendimento nominal mensal das pessoas responsáveis por domicílios particulares permanentes (com rendimento) (-) ▪ Valor do rendimento nominal médio mensal das pessoas de 10 anos ou mais de idade (com e sem rendimento) (+) ▪ Domicílios particulares permanentes do tipo casa com abastecimento de água da rede geral (+) ▪ Pessoas alfabetizadas com 80 anos ou mais de idade (+)
SÃO CAETANO DO SUL (SP)	-
SÃO PAULO (SP)	-
SERRA AZUL (SP)	-
VITÓRIA (ES)	-

Fonte: Própria Autora.